Development of Novel Approaches Towards the Study of Ribosome Bound Nascent Chain Complexes by NMR Spectroscopy

by

Jack Fred Sidney Paton

Submitted to the Institute of Structural and Molecular Biology (ISMB) for the requirements for the degree of Doctor of Philosophy

At the

Institute of Structural and Molecular Biology
University College London
London, WC1E 6BT, United Kingdom

September 2016

Supported by
wellcome trust

Declaration

I, **Jack Paton**, declare that all the work presented in this thesis is the result of my work only. Where information has been derived from other sources, I confirm that this is clearly stated in the thesis. The work herein was carried out while I was a graduate student at the University College London, Research Department of Structural and Molecular Biology, under the supervision of Professor John Chistodoulou.

Abstract

Co-translational folding is a fundamental process in biology studied primarily using biochemical and biophysical techniques. The dynamic nascent chain has largely eluded high resolution structural study, although NMR spectroscopy is recently emerging as one of the principal techniques for acquiring a dynamic and structural understanding of the nascent chain due its advantages in investigating dynamic systems. As such, NMR has lead to a rich understanding of ribosome-bound nascent chains (RNCs) in terms of their dynamics and folding on the ribosome.

The aim of this project is to explore the NMR observable properties that can be manipulated according to the kind of system one chooses to study. In particular, the exploitation of the fast exchange regime in NMR spectroscopy has not been applied to the study of RNCs. This is despite the potential advantages fast exchange can show over slow exchange for samples of low concentration and stability, which make it potentially very advantageous for the study of RNCs.

We briefly describe the previous approaches that have been successful in producing nascent chain samples that can be studied by NMR, namely α-synuclein, an intrinsically disordered protein (IDP), that allows for investigations of interactions with the ribosome surface, and domain 5 of the gelation factor of *Dictyostelium discoideum*, where co-translational folding has been probed by the development of constructs that possess varying degrees of structure whilst attached to the ribosome.

Following this, an argument is presented for the advantages in exploiting the fast exchange regime, as the domain 5 system is in slow exchange on the NMR timescale, which leads to certain limitations. This is followed by the design and development of translation stalled nascent chains of fast folding domains. We present NMR characterisation of the folding of two such fast folding proteins, the GA module and the HP36 villin headpiece domain. This is followed by NMR investigations into a translation stalled construct of the GA module, a small albumin binding domain.

Further to this, we also investigate the interaction between α-synuclein and the ribosome surface by measurement of amide proton relaxation rates of α-synuclein over a titration of varying α-syn concentrations. Finally, we show the results of attempts to incorporate 3-fluorotyrosine into proteins and RNC samples, and show how this can be used as a unique NMR probe. We show that this approach allows one to acquire NMR spectra of RNCs with no background signal from the ribosome, and that the sensitivity of the fluorine chemical shift to its environment therefore has the potential to uncover rich information about the ribosome surface and the effect it has on interacting species.

Results shown in this thesis were presented in the following papers:

Structural characterization of the interaction of α-synuclein nascent chains with the ribosomal surface and trigger factor. A. Deckert, C. A. Waudby, T. Wlodarski, A. S. Wentink, X. Wang, J. P. Kirkpatrick, J. F. S. Paton, C. Camilloni, P. Kukic, C. M. Dobson, M. Vendruscolo, L. D. Cabrita, and J. Christodoulou. Proceedings of the National Academy of Sciences, 113(18):5012-5017, 2016.

Results shown in this thesis were presented in the following meetings:

- **Poster presentations**
  - **Experimental Nuclear Magnetic Resonance (ENC) conference**, California, USA, 2012.
  - **UCL Graduate Symposia** 2013 and 2015.
- **Oral presentations**
  - **UCL Graduate Symposia** 2014 and 2016.
  - **CCPN conference** Buxton, UK, 2015.

Acknowledgements

Many thanks to my Supervisor, Prof. John Christodoulou, as well as Dr. Chris Waudby and Dr. Lisa Cabrita for their help and support and indispensable advice throughout this project.

My thanks also to the members of the Christodoulou group, in particular Dr Annika Deckert, for her guidance in my first days in the laboratory, Anaïs Cassaignau for providing previously collected data as well as guiding me through the process of collecting NMR data on nascent chains, Sammy Chan for many helpful discussions, and Dr Xiaolin Wang for providing the ^{1}H-^{15}N NMR spectrum of the L7/L12 domain of the ribosome.

I also send my thanks to Prof. Dan Raleigh and Dr Sarah Dodd (née Lejon) for their help in providing DNA constructs for this project, Dr Toshio Iwasaki for his kind donation of auxotrophic bacterial strains, Dr Mark Howard for his expertise in fluorine NMR, and Dr John Kirkpatrick for his extensive expertise over the many practical and theoretical aspects of performing NMR experiments and appropriately analysing data. Also to Dr Kersti Karu for her help in setting up mass spectrometry experiments.

For the lineshape simulations and the analysis of the α-synuclein titration data, my particular gratitude again towards Dr Chris Waudby and Dr John Kirkpatrick for detailed help with the theory and finding appropriate solutions to the Bloch-McConnell equations.

Gratitude also to my thesis committee, Dr Flemming Hansen and Dr Renos Savva for guiding me through the project at some crucial stages, and help with making some important decisions.

Contents

List of abbreviations

ΔTF	Trigger factor knockout
AAT	α1-antitrypsin
BEST	Band-selective Excitation Short Transient
CD	Circular Dichroism
Cryo-EM	Cryo-electron microscopy
DOSY	Diffusion Ordered Spectroscopy
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
ESI	Electrospray Ionisation
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMQC	Heteronuclear Multiple Quantum Coherence
HSQC	Heteronuclear Single Quantum Coherence
IDP	Intrinsically Disordered Protein
ILV	Isoleucine, Leucine, Valine
IPTG	Isopropyl β-D-1-thiogalactopyranoside
MBP	Maltose Binding Protein
NMR	Nuclear Magnetic Resonance
OD	Optical Density
PAGE	Poly-Acrylamide Gel Electrophoresis
PTC	Peptidyl Transferase Centre
RF	Radiofrequency
RNC	Ribosome-bound Nascent Chain
SDS	Sodium Dodecyl Sulphate
SOFAST	band-Selective Optimized Flip-Angle Short-Transient
SORDID	Signal Optimization with Recovery in Diffusion Delays
STE	Stimulated Echo
TEV	Tobacco Etch Virus
TF	Trigger Factor
TFA	Trifluoroacetic acid
TOCSY	Total Correlation Spectroscopy
TROSY	Transverse Relaxation Optimised Spectroscopy
VMD	Visual Molecular Dynamics
XSTE	Heteronuclear stimulated echo

List of figures

List of tables

‘Begin at the beginning’, the King said gravely, ‘and go on till you come to the end: then stop.’

-

Lewis Carroll, *Alice in Wonderland*

Chapter 1

Introduction

1.1 Protein folding on the ribosome

An understanding of the differences between the folding of an isolated protein and that of the same molecule undergoing translation is a crucial step for understanding protein folding in a biological context. The ribosomal complex is known to have an impact on the translating polypeptide as it moves away from the peptidyl transferase centre (PTC) and along the ribosomal exit tunnel (Figure 1.1), and has even been shown to interact with the partially folded protein as it emerges into the cellular milieu [2, 3]. These interactions are likely to play a role in modulating the folding pathway of a nascent chain, and numerous techniques have been developed to study the acquisition of structure on the ribosome. By using functional assays [4] it was possible to deduce that the ribosome has an impact upon the kinetics of folding, as it was observed that the bacterial luciferase enzyme becomes enzymatically active more quickly after cell free translation than after renaturation *in vitro* from urea induced unfolding. Single molecule force microscopy [2] has also been used to probe the effect of the ribosome on folding rates. As well as this, conformation dependant antibody recognition studies [5] indicated that different proteins acquire structure on the ribosome at different points in translation, and fluorescence anisotropy [6] has indicated unique dynamic behaviour of emergent peptides, and that such dynamic behaviour is characteristic of folding competent sequences. Cryoelectron microscopy [7] has revealed the presence of interactions between the nascent peptide and the walls of the ribosome exit tunnel, which have been inferred to be important regulators of translation rate, as well as having an effect on the conformation of the peptide.

Alongside these approaches, NMR spectroscopy has also emerged as a powerful technique for investigating

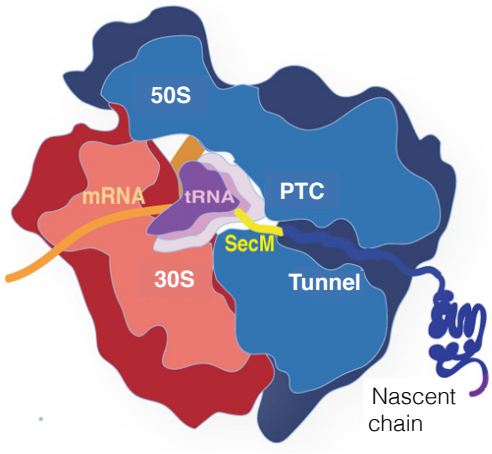

Figure 1.1: *Schematic diagram of a nascent chain complex.* After biosynthesis at the peptidyl transferase centre, the newly formed peptide chain begins the process of traversing from the core of the ribosome, through the tunnel, and into the cytosol, where it can begin the folding process. Figure obtained directly from Cabrita et al. [1].

this process, as it is capable of providing structural and dynamical details on the acquistion of structure of emerging nascent chains [8–11], which has been bolstered by developments in NMR methodology such as SOFAST and TROSY based experiments in conjunction with selective methyl labelling strategies [12, 13]. Particularly notable in the case of NMR are the extensive studies carried out on domain 5 of the gelation factor protein from *Dictyostelium discoideum* (FLN750) [11], which have shown that the fully translated domain exists in a folded state even whilst attached to the ribosome, showing that folding can occur before release of the protein from the PTC. Detailed analysis of selected heteronuclear multiple quantum coherence (HMQC) peaks of the nascent chain also revealed the presence of interactions between the N-terminus of the partially folded protein and the ribosome, thus displaying the unique power of NMR to provide structural and dynamic information at a residue specific level [11, 14].

In the following work, we discuss how previous research has uncovered the role played by the ribosome in mediating the folding process. The work carried out on the previous nascent chain systems is described, and certain limitations are pointed out. We then develop an argument to circumvent these limitations by appropriate choice of an alternate protein system with different NMR observable properties. We argue that the manner in which chemical exchange kinetics alter the appearance of NMR spectra is a critical component of choosing the appropriate system to maximise chances of developing a system that provides as much information as possible. We begin in the next section by outlining the fundamental nature of understanding how proteins fold in a general context, before applying this to a nascent chain system.

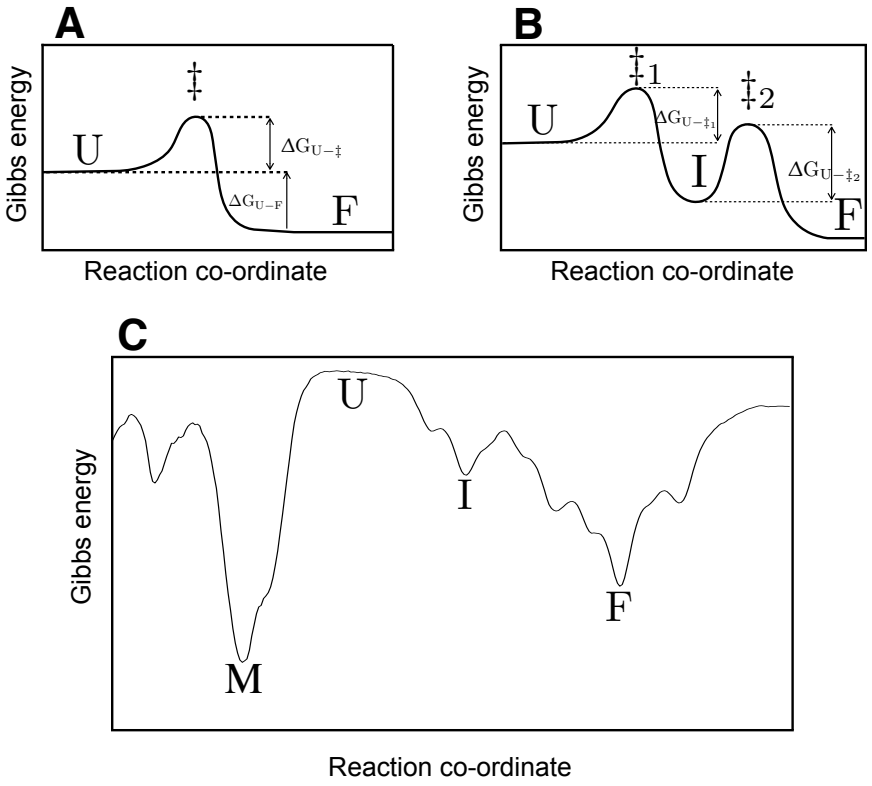

Figure 1.2: *Example protein folding energy landscapes.* **A**: Energy landscape for a two state folding system (equation 1.2.1). U = unfolded protein, F = native state (folded), ‡ = transition state. The kinetics of folding is determined by the Gibbs energy of activation, $\Delta G_{U-‡}$. The difference in Gibbs energy between the initial and final states, ΔG_{U-F}, determines the relative population of denatured to native protein at thermal equilibrium. **B**: Energy landscape for a protein that folds in a three state manner (equation 1.2.2). A stable intermediate is denoted with I, and transition states/Gibbs energies are numbered according to their order of appearance along the reaction co-ordinate. **C**: Alternate energy landscape schematic of an isolated protein, showing the potential existence of kinetic traps such as the M state, which represents very low energy states that can be accessed from the natively unfolded protein. These can be aggregate formations, or simply alternate conformations that have no biological activity. This figure was produced using the keynote software that comes installed on the Mac operating system.

1.2 The protein folding problem

The protein folding problem can be considered as the question of how a protein effectively exchanges between its denatured, unfolded state, U, and the native state, F. In the simplest case, one can imagine two state chemical exchange;

$$F \rightleftharpoons U. \tag{1.2.1}$$

Or there may be an intermediate, I;

$$F \rightleftharpoons I \rightleftharpoons U \tag{1.2.2}$$

This process can be stated to be in chemical equilibrium when the overall Gibbs energy G of the system is minimised, where G can be defined in terms of the enthalpy H, temperature, T, and entropy, S, of the system;

$$G = H - TS \tag{1.2.3}$$

The overall difference in Gibbs energy between two states such as U and F, is the crucial determinant of relative population of these two states;

$$\Delta G_{U-F} = G_F - G_U \tag{1.2.4}$$

which can also be related to the equilibrium constant, K, between the two states;

$$\Delta G_{U-F} = -RT\ln(K) = -RT\ln\left(\frac{[F]}{[U]}\right) \tag{1.2.5}$$

where T is temperature and R is the molar gas constant. Figure 1.2A and 1.2B show the typical energy landscapes that are described by equations 1.2.1 and 1.2.2, respectively. This reveals also the presence of transition states (T), that reside at the peak of the energy barrier that must be overcome to transition from one minimum to another. This barrier, known as the activation energy, $\Delta G_{U-\ddagger}$, defines the kinetics of exchange. The greater the activation energy, the slower the rate of exchange.

Further to this simple representation of folding, it is often the case that there may be numerous intermediates along the folding pathway, as well as alternate low energy conformations that do not possess a biological function. This kind of landscape is shown in Figure 1.2C, where the unfolded (U) state is capable of folding either into the folded (F) state, or a kinetic trap, described here as the M (misfolded) state, which is thermodynamically stable, but has no biological function, or may even be harmful. This energy landscape approach to protein folding is the predominant theory used to explain how proteins acquire their tertiary structure due to their funnel like conformational landscape that guides the protein to its native state [15, 16]. However, the existence of folding chaperones would indicate that this model does not accurately explain how all proteins fold, as is explained in the following section.

1.3 Chaperones guide the co and post-translational folding of proteins

A number of biological mechanisms are employed by most cells to moderate protein folding and prevent access to mis-folded states. Typically in *E. coli*, after synthesis on the ribosome, the nascent chain comes into contact with trigger factor (TF) [17] before being released into the cytosol, where other chaperones such as Hsp70 [18] and the GroEL/ES chaperonin [19] cage are known to play a role in further chaperoning certain proteins to their fully folded forms.

The GroEL/GroES chaperonin system is often described as a protein folding cage. Various theories have been employed to explain how this system works, which essentially revolve around ensuring that translated proteins are able to adopt the correct fold and are prevented from populating undesired minima in the energy landscape. For example, by encaging the protein, it has been postulated that the cage prevents the unfolded protein from interacting with other species and forming aggregates. Hence, it is only released after folding is achieved, and aggregation is no longer possible. This is known as the passive cage model [19, 20]. Alternately, in the iterative annealing model, the cage has been proposed to be capable of unfolding misfolded proteins and refolding them in an iterative cycle [21, 22], making it a more active participant in the folding process. In this case, one can imagine a scenario such as in Figure 1.2C, where the caged protein is not released when it adopts the M state, but instead is unfolded and refolded again, until the F state is reached, which then triggers release of the protein. In this way, the polypeptide is deliberately guided down the alternate pathway, even if the M state is more thermodynamically stable for the isolated protein.

The chaperoning of folding can also occur whilst the protein is being translated. One of the most well studied chaperones involved in this process is the trigger factor (TF) protein, which forms a complex at the exit tunnel of the ribosome, allowing it to act as a 'cradle' for translating proteins [17, 23–26].

By interacting directly with the nascent peptide as it emerges from the tunnel, this 48 kDa peptide is known to directly reduce the rate of protein folding on the ribosome, whilst simultaneously shielding it from unwanted non-native interactions with other species within the cytosol of the cell [27, 28]. Only once a sufficient length of the peptide has been translated is the trigger factor thought to allow folding to occur [29]. This has been proposed to be crucial in preventing aggregation of newly translated proteins [30]. The importance of trigger factor in co-translational folding has been confirmed for maltose binding protein (MBP) by optical tweezer experiments, where it was shown to stabilise on-pathway folding intermediates [31], and there are many other examples of the impact of trigger factor upon the folding of partially translated proteins

at the ribosome exit tunnel [32, 33]. The importance of these and other chaperones for protein folding is indicative of the notion that a protein often needs some form of guidance within the cell in order to ensure that it forms the correct fold, which is somewhat opposed to Anfinsen's dogma [34, 35], which in some forms states that the primary sequence alone is sufficient for a protein to acquire its native tertiary fold. We intend to explore further the role that the ribosome might play in chaperoning a protein, given that it is the first species that a new protein comes into contact with.

A question of exceptional interest in biology is that of the role the ribosome plays in modulating energy landscapes such as those shown in Figure 1.2C [36–38]. Although it has been demonstrated that the ribosome plays a role in altering the kinetics and thermodynamics of folding, the precise nature of this role, and its relevance to protein folding, is not fully understood. A simple means of gaining insight into how the ribosome affects folding of a protein is to consider how thermodynamic aspects of a chemical exchange process can be applied to a co-translational environment. For example, considering equation 1.2.5, it is clear that the presence of the ribosome affects the value of ΔG_{U-F} by altering the energy of either the native or non-native states, hence immediately affecting the equilibrium. If one can determine the impact that the ribosome has on the relative populations of folded and unfolded states on the ribosome, this is direct insight into the impact upon these energies.

Figure 1.3 shows a schematic visualisation of the pathway we are interested in. The coupling between the acquisition of tertiary structure and translation of a protein leads to a scenario where the conformational space available to the polypeptide is constantly changing as it emerges from the ribosomal exit tunnel, and certain misfolded states may also be unavailable as a consequence of the ribosomal environment. By determining how the exchange equilibrium of folding on the ribosome is affected at multiple points in the translation process, we obtain a series of snapshots, with altering thermodynamic equilibria being used as a monitor of the impact of the ribosome on co-translational acquisition of structure. We can then build a picture of how the landscape of folding is changing as the protein emerges from the tunnel. This constant alteration of the conformational energy landscape of the protein as it is translated is an aspect of co-translational folding that must be explored if we are to understand how a polypeptide acquires its structure in a biological context.

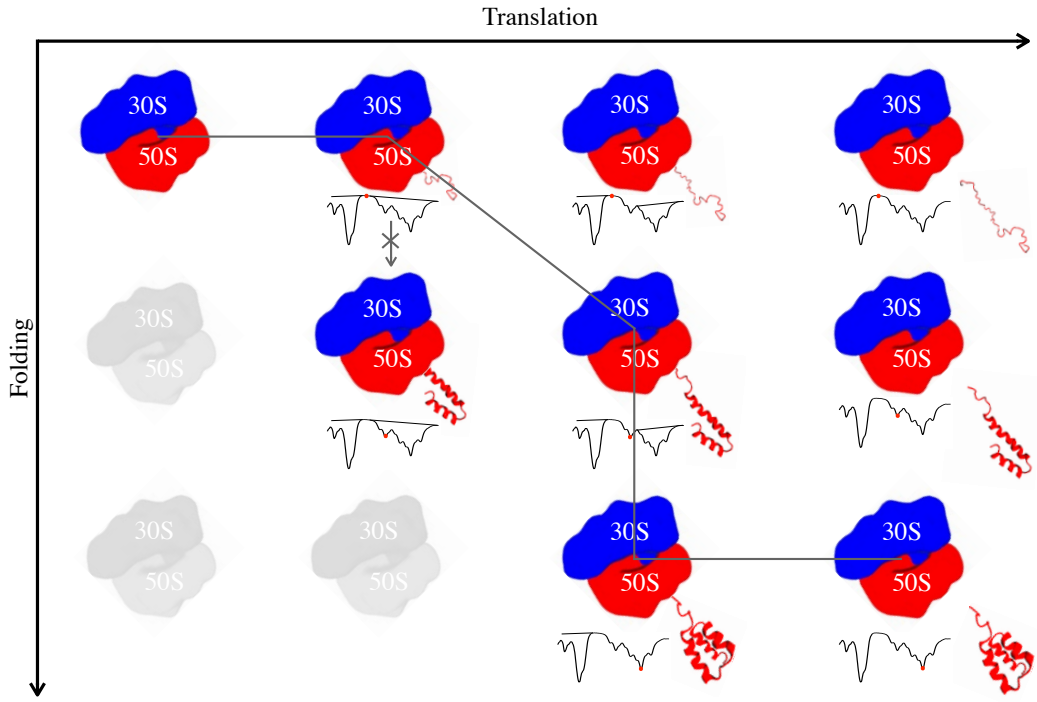

Figure 1.3: *The Coupling of Folding with Translation.* Schematic diagram showing the co-translational acquisition of structure of a nascent polypeptide chain, coupled with its emergence from the ribosomal tunnel during biosynthesis. The grey line indicates a hypothetical pathway that may be taken by a protein with comparable folding and translation rates. Under each nascent chain complex is a representation of the energy landscape from figure 1.2C. The red dot indicates which point of the landscape is occupied by the nascent chain for that complex, and the straight lines indicate how the ribosome is expected to modulate the conformational landscape, making certain areas energetically inaccessible. This allows the ribosome to funnel the translating peptide towards the folded state whilst preventing the population of undesirable, misfolded states that may be accessible to the unfolded, isolated protein (such as in the top right complex). This figure was produced using the keynote software that comes installed on the Mac operating system.

1.3.1 The nature of the ribosomal environment

The ribosomal exit tunnel forms the path taken by all proteins after they are formed at the PTC, hence leading from the core of the complex to the cytosol. It is known to be formed mostly of RNA, with some small protrusions of ribosomal proteins into the tunnel, namely L4 and L22 [39]. It is useful to consider how the various properties of this tunnel may influence folding. In a simplistic sense, the ribosome may have an inhibitory impact on folding purely as a result of sterics, whereby the exit tunnel cannot accomodate tertiary structure purely because it is too narrow. Simulations have been carried out which explore the steric effects

of the exit tunnel in some detail [40, 41], suggesting that the ribosome can be considered an inhibitor of folding at least on steric grounds.

However, one must also consider other properties of the tunnel that contribute to the overall environment. The tunnel is composed of hydrophobic, hydrophilic, polar, non-polar, positively and negatively charged regions, and the impact of these upon the behaviour of any translating protein is almost certain to be complex and computationally challenging. There is however, a large experimental body of information available describing the structural details of polypeptides within the tunnel. Medium sized proteins have been shown to be capable of forming tertiary structures near the exit of the ribosomal tunnel [41], and small protein domains have even been shown to be capable of folding within the tunnel [42]. Interactions between the translating peptide and the tunnel have been detected or can be inferred from a wide number of studies [2, 7, 11, 43–45]. Also, with the use of NMR and fluorescence techniques [6] we are able to explore the dynamic protein outside the exit tunnel, and how it interacts with the ribosome surface. This has revealed that electrostatic factors play a large role in modulating the dynamic behaviour of nascent chains [3]. The largely negative surface of the ribosome seems able to attract, and restrict the dynamics of, positively charged nascent chains, while its repulsive effect on negatively charged proteins causes such species to populate more dynamic conformations.

The roles played by the interactions between translating proteins and the ribosome vestibule, exit tunnel or surface are varied. For example, it may be that these interactions modulate the folding landscape of proteins to guide them towards the correct fold. Another important role might be to control the rate of translation, allowing regulation of biosynthesis [46], and in some cases inducing translational stalling, allowing regulation of gene expression [7]. It is therefore very important to understand how these interactions affect the conformation and translation rate of a protein.

To understand how such interactions may influence the folding behaviour of a protein, simulations by O'Brien et al. [47] show that when a polypetide is confined within a carbon nanotube, the alpha-helical content of helix forming peptides is governed not just by steric factors (i.e. tunnel width), but also by the presence or absence of hydrophobic interactions between the peptide and the wall of the tube. Consider, for example, the amphiphilic poly-AS sequence studied in [47]. On formation of an α-helix, amphiphilic peptides display hydrophobicity on one side, and hydrophilicity on the other side of the helical coil. If there is a tendency for α-helical peptides to adsorb to the surface of the nanotube (or tunnel), this adsorption will help to enthalpically stabilise the secondary structure as the hydrophobic patch on one side of the helix binds preferentially to the surface. The simulated carbon nanotube therefore represents just one factor that

is present within the tunnel, namely a hydrophobic surface which could potentially influence the formation of helices. This does not suggest that this is the co-translational mechanism of α-helix formation. Rather, it indicates that the interplay between the physical properties of the translating protein and the immediate surface of the inside of the ribosome tunnel can strongly influence preferred conformations of the protein. Other factors, such as electrostatics, will also be simultaneously affecting conformational behaviour, making the co-translational environment markedly different from the typical *in vitro* conditions that folding is usually studied under.

Moreover, the hydrophobic, electrostatic, steric and other factors that influence this folding process are also undergoing a constant process of changing levels of prominence as the protein translates. Hence, for every new amino acid, the protein is driven towards a new pseudo-equilibrium state. This indicates that the exit tunnel cannot be considered merely as a static, inert channel through which the protein emerges with no significant interactions, but rather, it features a rich range of chemical moieties and dynamic, fluctuating surfaces that have complex interactions with passing polypeptide chains. One might argue that the impact of the ribosome on folding is less important because any given protein is only present at the ribosome surface for a limited time before it is transported elsewhere. If the kinetics of translation are much faster than the kinetics of folding, then the ribosome may not influence folding to any great degree at all. However, the rate of translation for an *E. coli* ribosome is known to be in the range of 12-21 amino acids per second [48]. Hence, each time a new amino acid is added to a translating protein, there is a period of between 50 and 80 milliseconds before the next amino acid is added to the chain. A protein of 100 amino acids would take approximately 5 to 8 seconds to become fully translated. Since many proteins fold on timescales much faster than seconds, one can posit that their folding behaviour is meaningfully affected by the ribosome surface. We now describe the strategies and approaches that can be used to isolate high quality samples of nascent chain complexes that can be investigated using NMR spectroscopy.

1.4 Strategies for production of nascent chains suitable for NMR

One of the core experimental techniques necessary to this project is the design and production of a nascent chain sample. This design strategy can be broken down into a number of separate considerations; An appropriate protein must be identified which promises to yield new information about folding on the ribosome. As will be discussed below, kinetics is an important consideration here, since inappropriate folding kinetics can render the protein unobservable by NMR, and hence no information can be obtained. In a similar

vein, one must also consider whether interactions with the ribosomes might also lead to line-broadening that renders the protein impossible to observe.

Once a protein has been identified as viable for study, it must be cloned into an expression vector that allows for homogeneous stalling of expression on the ribosome. This vector typically consists of a cleavable N-terminal polyhistidine purification tag and the inclusion at the C-terminus of the SecM sequence [11]. This 17 amino acid stretch is known to stall translation by physcially tethering the nascent chain to the inside of the exit tunnel, thus preventing release from the tRNA [45, 49]. The SecM sequence and the protein of interest can be separated further by introduction of a linker sequence to probe the effects of increasing distance between translated protein and the PTC.

RNC samples are notable for their instability, and their low maximum concentration. Typically, ribosome concentrations above 10 μM rapidly break down, and below this a sample can last for several days if produced to high purity and kept at low temperatures, (ca. 4 °C). However, typical conditions for acquisition of NMR data often tend to require 25 °C, which means some samples only last for a day, or a few hours, before degradation occurs. The process of sample break down is not fully understood, but it is known that nascent chain release, aggregation and ribosome decomposition are all contributing factors.

Another problem is the size of the complexes being studied. The majority of the 2.3 MDa ribosome cannot be observed by NMR due to its large rotational correlation time leading to extremely broad line-widths (see section 1.6.1). As such, only dynamic regions such as the L7/L12 stalk, with higher local correlation times than the average observed in the ribosome's core, can be seen by NMR [50]. Similarly, nascent chains can be observed by NMR provided they are sufficiently outside the ribosomal tunnel such that their movement is not restricted. However, in some cases it is still necessary to use sophisticated labelling techniques in conjunction with TROSY based pulse sequences in order to observe the required signal (see section 1.5 and 1.6.2).

Although rich in information, the low sensitivity of NMR spectroscopy tends to necessitate high sample concentrations compared with other techniques, typically ranging from 100 μM to several mM for isolated proteins. Many experiments, such as relaxation measurements, also require long durations of time, even at these high concentrations, and so they cannot be applied to a nascent chain which has a maximum concentration of 10 μM and is liable to decompose over the duration of the experiment. Much work has gone into developing NMR experiments that reduce overall acquisition time [51]. There have also been attempts to improve the sensitivity of samples by addition of components such as paramagnetic compounds that allow a greater number of scans to be carried out in a given time frame [52] be reducing longitudinal relaxation

times.

Another core aspect of the nascent chain production strategy is to ensure that the integrity of the sample can be preserved and also confirmed for as long as possible. This allows us to be sure that the sample under investigation actually corresponds to a translation-stalled protein. Once a nascent chain sample has been purified and is ready for NMR, it is therefore important to constantly monitor its status. The first and most critical strategy is the western immunoblot. By utilising antibodies for the SecM stalling sequence and the hexahistidine purification tag (two near-universal features of RNCs), expression of the nascent chain can be easily confirmed. This strategy is useful because an intact, stalled nascent chain is covalently bound to the last transfer RNA (tRNA) molecule at the PTC. If the sample has broken down, the nascent chain is released from the tRNA (of approximately 20 kDa) and this large change in molecular weight is a very simple monitor of the integrity of the sample (see Section 4.5.4 for experimental examples).

1.5 Isotopic labelling strategies for nascent chains

As well as ensuring the nascent chain sample has been produced to a high level of purity, we must also ensure that an appropriate isotopic labelling strategy has been employed to ensure that meaningful NMR data can be obtained from the sample. For the purposes of NMR, it is important to maximise the isotopic labelling of the nascent chain while minimising labelling of the ribosome background. Ideally, the ribosome should be completely unlabelled, or labelled with deuterium, an NMR silent nucleus, in order to completely eliminate background signal. In order to produce such samples, an expression method has been developed over the years by the Christodoulou lab that allows the selective labelling of a nascent chain with the desired isotopic components, and this approach has been used in much of the recent work carried out by this lab [11, 53]. Initially, cells are grown to saturation in a non-inducing medium before transferral to the induction medium. At this stage, the desired isotopes are added along with the induction agent isopropyl β-D-1-thiogalactopyranoside (IPTG) and the antibiotic rifampicin. The rifampicin inhibits ribosome growth, a goal which is also achieved by ensuring that the growth medium is saturated with *E. coli*. As a result, ribosome growth is minimised during the period of expression, preventing them from incorporating the isotopes intended for the nascent chain. This allows us to maximise the NMR observable signal that comes from the nascent chain (see section 7.5 for further details). What is most important, and we consider this here, is the choice of which isotopes are introduced into the nascent chain and/or the ribosome in order that as much information can be obtained as possible about the system.

The chemical shift of ^{15}N in the amide backbone of a protein chain is strongly dependant upon primary sequence. Although secondary and tertiary structure also impacts the resonant frequency, this primary sequence dependance leads to large dispersion of ^{15}N chemical shifts even in a fully unfolded protein. Therefore, ^{15}N labelling is commonly adopted as a means to investigate unfolded and disordered proteins, since it increases the possibility of following a single ^{1}H-^{15}N peak from folded to unfolded without interference from other resonances. For this reason, ^{15}N labelled nascent chains have been effectively used for the study of α-synuclein, an intrinsically disordered protein [53], as well as the unfolded domain 5 of ddFLN, [11]. However, the acquisition of structure can significantly reduce the rotational correlation time of backbone amide groups, since a folded protein tends to have a backbone far more rigid than its disordered counterpart. When this is combined with attachment to a massive body such as the ribosome, the line-broadening that results typically renders structured ^{1}H-^{15}N signal impossible to observe.

A successful approach to overcome the challenge of observing a structured nascent chain attached to the ribosome has been found to be the selective $^{13}CH_3$ labelling of methyl groups of isoleucine, leucine and valine (ILV) residues against a background of ^{2}H and ^{12}C. By only labelling selected methyls with ^{1}H, one reduces proton density and hence dipolar relaxation processes. The choice of methyl group is made because the symmetry of the CH_3 moiety means that one observes three protons as though they are a single entity, giving a stronger signal. This is due to rapid rotation of the methyl group about the C-C bond. We also observe the methyl TROSY effect with the CH_3 moiety; due to cross-correlated relaxation, certain coherences have narrow linewidths, and these can be selectively observed against coherences with broader linewidths for the same nucleus. The HMQC pulse sequence, which is used to collect methyl-HMQC spectra, is optimal for observation of such coherences.

The major drawbacks with this labelling strategy are firstly the monetary cost of sample production, as the selectively isotope labelled precursors and D_2O necessary for sample preparation are very expensive. Secondly, the poor dispersion of methyl peaks in unfolded states also limits the ability to observe all states of a folding process simultaneously. In a structured protein, the ILV methyls tend to have unique chemical shifts with little overlap, but on unfolding, they all collapse to nearly identical chemical shifts in the ^{1}H and ^{13}C dimensions. Therefore, this strategy can only be adopted for structured nascent chains. However, it has been very successful in providing residue specific details on the structure of folded, ribosome-bound nascent chains [11].

1.6 NMR studies of nascent chains and other macromolecular complexes

Described in this section is a summary of NMR studies of large macromolecular complexes carried out in recent years, as well as specific focus on the NMR of nascent chain complexes. We begin with a discussion of the problems associated with NMR of macromolecules, using the ribosome as an example of the advantages and disadvantages of this approach.

1.6.1 NMR of large species

As described in section 1.7.2, the rotational correlation time of a molecule in solution, τ_c, is strongly related to its molecular mass. Larger species tumble slowly, leading to a decrease in the transverse relaxation time, T_2. There are 2 main ways in which this problem can be alleviated. One can, for example, focus only on dynamic regions of large molecules. As well as this, the use of transverse relaxation optimised spectroscopy (TROSY) is also able to greatly increase the molecular size that can be observed by NMR.

Consider the *E. coli* ribosome. Initial structural studies on this complex were carried out in 2000 using x-ray crystallography [54–57], which allowed high resolution structural information to be gained on the complex. The level of structural detail that was gained for large sections of the ribosome, including the core and most proteins near the surface, cannot be matched by NMR spectroscopy as the molecular mass of the ribosome (2.3 MDa) leads to such slow tumbling that relaxation times are exceptionally short. However, dynamic features of the complex elude characterisation by crystallographic methods. In particular, the L7/12 stalk of the ribosome cannot be studied on the complex using either crystallography or cryo-electron microscopy. This extremely dynamic region of the ribosome has such flexibility that it does not occupy a single homogenous conformation even in a crystal. However, this dynamic behaviour renders it entirely observable by NMR, as was demonstrated in 2004 [50, 58]. The flexibility of this region leads to a higher local correlation time than the rest of the ribosome, and thus line-widths are far more narrow than would be expected for a 2.3 MDa molecule.

The GroEL/ES chaperonin system is another example of an extremely large species that has nevertheless been successfully characterised by NMR [59]. In this case, the use of transverse relaxation optimised spectroscopy (TROSY) based techniques were instrumental in gaining high quality spectra of this 800 kDa complex. The unparalleled ability of NMR to give dynamic and residue specific information alongside structure, not to mention its capacity for elucidating structures that cannot be accessed by other techniques (such

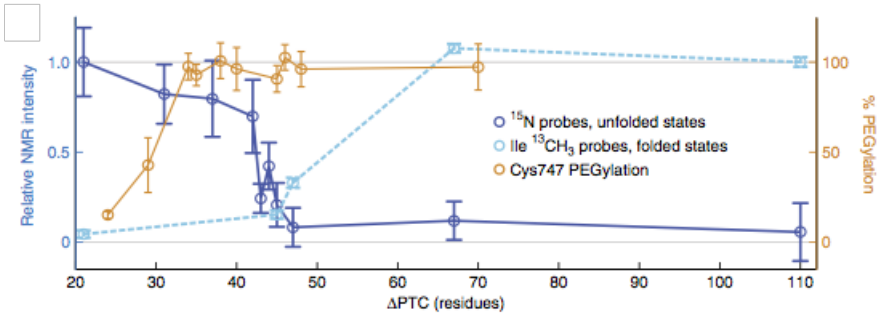

Figure 1.4: *Intensity plot of FLN750 RNC residues as a function of domain 6 linker length.* Folding of the FLN750 nascent chain, measured by intensities of ^{15}N amide resonances in blue and isoleucine $^{13}CH_3$ resonances in cyan as a function of linker length. Also shown in orange is the plot of the extent of PEGylation of the native Cys747, which is not relevant for discussion here. Figure was taken directly from [11].

as crystallography and cryo-EM), makes it an extremely powerful tool to complement these techniques. Also demonstrated here is the range of dynamic behaviours exhibited by protein structures. Even macromolecules in the megadalton range can be studied by NMR, indicating that there must be regions of high local dynamic behaviour, since NMR spectroscopy is uniquely suited to detecting such regions.

1.6.2 NMR spectroscopy of nascent chains

We now discuss the folding characterisation of some nascent chain complexes, namely domain 5 of the gelation factor protein from *Dictyostelium discoideum*, and α-synuclein. The domain 5 (or FLN5) protein constitutes residues 646 to 750 of the gelation factor, and so, from this point, all references to the FLN domain 5 protein will term the full protein as domain 5, or FLN750, in order to distinguish it later on from the C-terminally truncated variant, which lacks 12 amino acids and is referred to as FLN738.

The FLN750 domain has been extensively studied as an isolated protein by NMR, and its structure and dynamical and folding behaviour is well understood [60]. Over the years, it has also been the subject of very detailed research as a model nascent chain [1, 11, 12, 14]. We will focus more on the latest publication at the time of writing [11], because it contains the most relevant information to the topic at hand.

A range of constructs have been developed, with linker lengths ranging in length from +21 to +110, where +21 indicates 21 amino acids between the PTC and the beginning of the C-terminus of domain 5. Two isotopic labelling schemes were used to investigate these constructs. A uniform ^{15}N labelling strategy was utilised to provide maximum dispersion in observing the unfolded domain, and information was gathered on this sample using a SOFAST-HMQC based pulse sequence [61]. Selective ^{13}C-^{1}H labelling of isoleucine

methyls against a ^{2}H-^{12}C background was able to identify the tertiary structure that could be observed at longer linker lengths, due to the aforementioned favourable relaxation properties of this labelling scheme. These samples were investigated using a methyl TROSY pulse sequence [62].

Resonance intensities were mapped as a function of linker length, or ΔPTC, in order to make estimates of the relative population of folded and unfolded states (See Figure 1.4). What is clear from this data is that as linker length increases, the intensity of unfolded peaks is decreased as the intensity of the folded peaks increases, suggesting a gradual acquisition of structure during translation. This information could be gained entirely from a careful evaluation of peak intensity and line-broadening, highlighting the power of the NMR technique.

The α-syn nascent chain

Where the FLN750 system was an effective probe of folding on the ribosome, studies carried out on α-syn as it emerges from the ribosome have been instrumental in detailing interactions between disordered peptide and ribosome surface, as well as the chaperone, trigger factor. We will briefly discuss results from the work by Deckert et al. to illustrate this point [53].

NMR experiments on isolated α-syn in the presence of 70S ribosomes have revealed line-broadening in ^{1}H-^{15}N HSQC/HMQC peak intensities, indicative of an interaction between the isolated IDP and the 70S complex. Since α-syn is a disordered protein, this was an opportunity to probe the nature of interactions between a nascent polypeptide and the ribosome, since NMR observations of the α-syn nascent chain can be attributed to interactions, as opposed to folding events.

Hence, the α-syn nascent chain was produced with uniform ^{15}N labelling against an unlabelled ribosome background in an analogous manner to FLN750. However there was no linker between the SecM and the nascent peptide. Instead of varying linker length, charge mutations were inserted into the wild type, in order to evaluate the impact of electrostatics. It was found that introduction of reverse charge mutations, replacing positive with negatively charged residues, leads to increased amide peak intensity, reflecting the increased dynamic behaviour of the nascent chain, as it became repelled to a greater extent by the negatively charged ribosome surface.

Further experiments also found that trigger factor binds to the protruding α-syn sequence whilst it is attached to the ribosome, indicating that it plays a role in chaperoning the unfolded protein as it emerges into the cellular milieu. This displays the power of NMR in providing a great level of structural and dynamic detail on a complex molecular process.

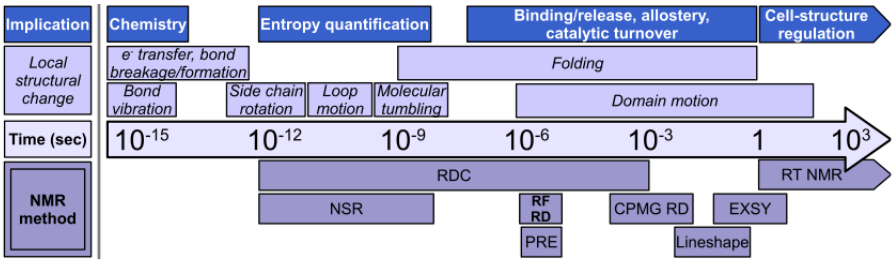

Figure 1.5: *Timescale range NMR techniques for study of protein dynamics.* Protein dynamics cover a very broad range of timescales from femtoseconds to seconds, all of which enable biological function in some form. As such, the ability of NMR to study a broad range of dynamic timescales makes it an extremely valuable tool. Figure copied directly from [66].

In the following section, we will describe some key background aspects to the use of NMR as a biophysical technique for the study of protein structure, dynamics and folding.

1.7 NMR background

1.7.1 Using NMR to Study Dynamic Systems and Chemical Exchange

Much of the theory described in this section is described in great detail in the book by Cavanagh et al. [63]. NMR spectroscopy can be used to study a broad range of dynamic time scales, from picosescond to millisecond and greater, given appropriate sets of conditions with suitable pulse sequences. Figure 1.5 shows the range of timescales accessible to various NMR experiments [64]. While this range is ostensibly very broad, there are caveats for each timescale. For example, CPMG relaxation dispersion techniques can be used to investigate dynamic processes on a millisecond timescale [65], however they typically require high concentrations and large amounts of time, due to the insensitivity of these types of experiments.

In the context of studying a large macromolecular complex of limited stability, such as an RNC, this puts many limitations upon the available experiments. Even with the labelling strategies mentioned in section 1.5, we can only reliably record 1D, 2D SOFAST and diffusion experiments. Relaxation experiments, NOEs, and more advanced 2D or 3D spectra all take so much time that sample breakdown over the course of acquisition would render the information gained from such experiments unreliable, and even then the signal gained from such low concentrations would be very weak. As well as limitations due to experimental time, chemical exchange processes can sometimes affect the appearance of NMR spectra in a way that renders the extraction of useful information extremely challenging, or impossible. To illustrate this point, we describe below how chemical exchange processes can affect the appearance of NMR peaks in a typical spectrum.

Consider a nucleus undergoing exchange between two chemical environments, A and B,

$$A \underset{k_{BA}}{\overset{k_{AB}}{\rightleftharpoons}} B \tag{1.7.1}$$

where k_{AB} and k_{BA} are the rates of the forward and reverse processes, respectively. The overall exchange rate, k_{ex}, is given by

$$k_{ex} = k_{AB} + k_{BA}. \tag{1.7.2}$$

The larmor frequency ν (s^{-1}) of the spin in the A state in the presence of a magnetic field of strength B_A is given by

$$\nu_A = \frac{\gamma B_0}{2\pi}\left(1 - 10^{-6}\delta_A\right) \tag{1.7.3}$$

where γ is the gyromagnetic ratio of the nucleus. The difference in Larmor frequency between the two exchanging states is

$$\Delta\nu = \nu_A - \nu_B. \tag{1.7.4}$$

The chemical shift (δ) is related to the Larmor frequency by the following equation;

$$\delta = 10^6 \times \frac{\nu - \nu_{ref}}{\nu_{ref}} \tag{1.7.5}$$

where ν_{ref} is a predetermined reference frequency. For biological samples, the reference compound is typically 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS).

The frequency of a peak can be represented as the angular frequency, ω, the Larmor frequency, ν, or the chemical shift, δ. Their definitions relative to each other are given thus;

$$\omega = 2\pi\nu \tag{1.7.6}$$

$$\delta = 10^6 \times \frac{\nu - \nu_{ref}}{\nu_{ref}} \tag{1.7.7}$$

We now consider the difference in angular frequency when compared against k_{ex}. This determines the exchange regime, and hence the appearance of an NMR spectrum. Using this model, we can describe three major scenarios of slow, fast and intermediate exchange, described below;

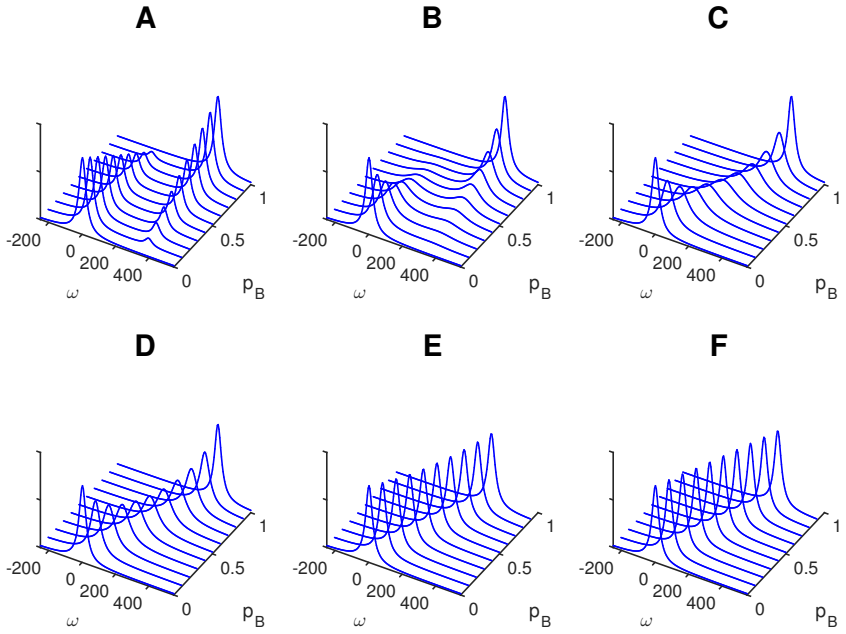

Figure 1.6: *Appearance of 1D NMR spectra for a species in chemical exchange under different exchange regimes.* Representation of the appearance of 1D NMR spectra for a precessing spin undergoing exchange according to equation 1.7.1, with the k_{ex}:$\Delta\omega$ ratio gradually increasing from **A** to **F**. One observes a transition from slow exchange, through intermediate and into fast exchange, with the corresponding change in spectral lineshape. **A:** $\frac{k_{ex}}{\Delta_\omega} = \frac{1}{350} \approx 0.003$ **B:** $\frac{k_{ex}}{\Delta_\omega} = \frac{100}{350} \approx 0.286$ **C:** $\frac{k_{ex}}{\Delta_\omega} = \frac{500}{350} \approx 1.429$ **D:** $\frac{k_{ex}}{\Delta_\omega} = \frac{1000}{350} \approx 2.857$ **E:** $\frac{k_{ex}}{\Delta_\omega} = \frac{10^4}{350} \approx 28.57$ **F:** $\frac{k_{ex}}{\Delta_\omega} = \frac{10^5}{350} \approx 285.7$.
Figure was produced using a matlab script to simulate and plot spectra [67]. The method used to simulate the spectra is similar to that shown in Appendix A.

$$k_{ex} \ll \Delta\omega \qquad \text{slow exchange} \tag{1.7.8a}$$

$$k_{ex} \approx \Delta\omega \qquad \text{intermediate exchange} \tag{1.7.8b}$$

$$k_{ex} \gg \Delta\omega \qquad \text{fast exchange} \tag{1.7.8c}$$

These scenarios all give rise to very different spectra, as outlined in Figure 1.6. In slow exchange, one observes two individual peaks at the larmor frequency of A and B, and their relative populations are given by the ratio of the peak height. In fast exchange, one observes a single peak, and the precise frequency at which this peak occurs, weighted to the frequency of the pure A and B states, determines the position of equilibrium. Intermediate exchange gives rise to a combination of line-broadening and change in peak

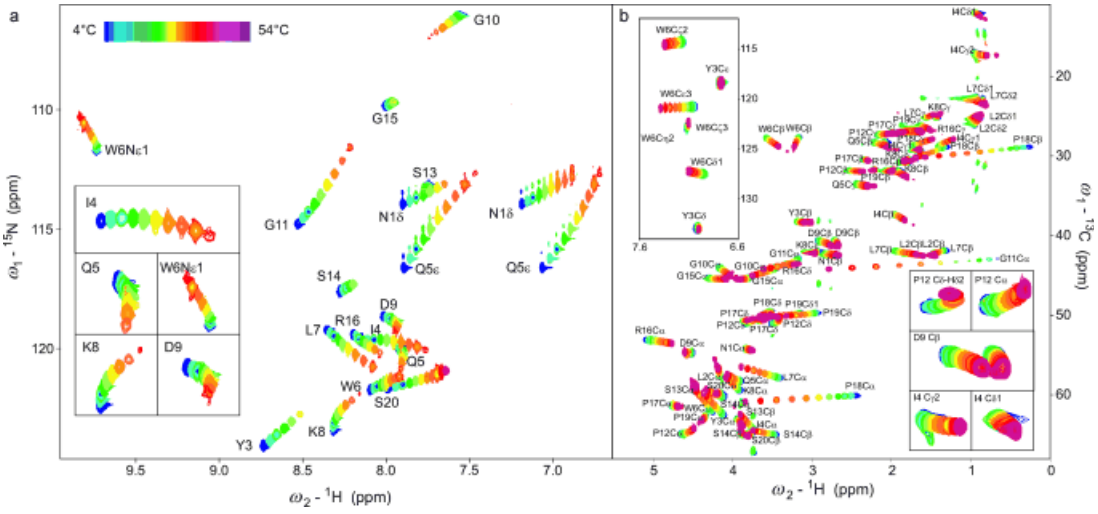

Figure 1.7: *^{1}H-^{15}N HSQC spectra of the Trp-cage protein at incremental temperatures.* The smooth transition in peak chemical shifts is a clear sign of fast exchange on the chemical shift timescale. Figure obtained directly from Rovo et al., [68].

position, and this can be problematic, as the more complex lineshape that arises can be more difficult to analyse without appropriate theoretical treatment. As well as this, peak intensity is lowest in the mid-range of the exchange process, which is often where the most interesting and useful information lies about the system being investigated. Hence, if the signal is too weak to be observed above the noise, it is difficult to meaningfully analyse this kind of exchange regime.

A practical example of fast exchange is given in Figure 1.7, (obtained from [68]), showing ^{1}H-^{15}N and ^{1}H-^{13}C correlation spectra of the Trp-cage miniprotein at various temperatures. The Trp-cage is a 20 residue peptide with a folding time, τ_f of approximately 4 µs, or a folding rate k_f of 2.5×10^5 s^{-1} [69] at 22.7 °C. In this case, one can see that as the protein was denatured by heating from 4 to 54 °C, the HSQC cross-peaks smoothly alter their position, indicating a gradual increase in the population of the denatured state at higher temperature. In this case, the A and B states of equation 1.7.1 correspond to the fully folded and disordered protein, respectively.

1.7.2 Longitudinal and transverse relaxation

Consider a nucleus held in a magnetic field B_0, which is thus aligned parallel with this magnetic field along the z axis. Application of a radiofrequency (RF) pulse to this nucleus in the xy plane will rotate this

alignment. If the correct power and duration of pulse is used, then the alignment can be rotated 90°, placing the magnetisation in the xy, or transverse, plane. The spin then precesses about the z-axis, whilst returning the equilibrium state of magnetisation, aligned again with the B_0 field.

Longitudinal relaxation is the return of magnetisation to the z axis after perturbation of magnetisation. The rate of this relaxation, R_1 (or relaxation time T_1), is defined by equation 1.7.9;

$$M_z(t) = M_{z,eq} - [M_{z,eq} - M_z(0)] \exp -R_1 t \tag{1.7.9}$$

where $M_z(t)$ is the longitudinal magnetisation at time t, $M_z(0)$ is the magnetisation at time 0 (immediately after the pulse), and $M_{z,eq}$ is longitudinal magnetisation at thermodynamic equilibrium.

Transverse relaxation is the loss of magnetisation from the transverse plane immediately after the radiofrequency pulse, and has a rate defined as R_2 (or relaxation time T_2);

$$M_{xy}(t) = M_{xy}(0) \exp -R_2 t. \tag{1.7.10}$$

As well as the natural propensity of the system to return to equilibrium, there is another factor that contributes to transverse relaxation, namely, the inhomogeneity of chemical shifts. As different spins precess at slightly different frequencies in the transverse plane, they move away from each other, eventually becoming distributed entirely about the transverse plane. At this, point, as their respective magnetisations point in opposite directions, the net magnetisation is close to zero, even though the overall magnetisation is still in the xy plane. Rapid transverse relaxation manifests itself in NMR spectra as broad lines. The faster the relaxation rate, the broader the line-width. It is also the case that larger molecules experience more rapid transverse relaxation as a result of their slower rotational correlation time, τ_c. Since the chemical shift of a particular nucleus depends upon its orientation in the magnetic field, at any point in time the chemical shift of all the nuclei in a specific position in a molecule has a broad range of values. If the molecules are tumbling rapidly, these frequencies collapse into a narrow, time averaged frequency. However, if tumbling is slow, the different range of precessional frequencies contributes to the transverse relaxation rate, and the lines become broader.

1.8 NMR experiments used in the present report

We now describe the NMR experiments used in the present report.

BEST sequence for collecting assignments

The Band-selective Excitation Short Transient (BEST) method is designed for the rapid acquisition of high quality multidimensional spectra such as is used for assignments. It is useful for low concentration samples, or in the case where one has limited time in which to collect a set of spectra in which to carry out an assignment. It works by using band-selective pulses to only excite the protons of interest (such as backbone amide protons). This prevents excitation of ^{1}H nuclei outside the selected frequency range, the main consequence of this being that these unperturbed spins cause rapid dipolar relaxation of the perturbed spins. The increase in the overall longitudinal relaxation rate R_1 (relaxation time T_1) allows for short repetition delays, allowing more spectra to be collected in a given time frame [70, 71].

Measurement of longitudinal relaxation (T_1) by inversion recovery

A typical means of measuring longitudinal (T_1) relaxation is to excite the spin of interest with a 180° pulse. For a sample at equilibrium, this aligns magnetisation directly against the applied magnetic field, and is described as being in the -z axis, whereupon it begins to relax back to +z at the rate R_1 ($= \frac{1}{T_1}$) . After a delay, t, the sample is then irradiated with a 90° pulse and the remaining signal observed. This signal depends upon the delay t according to equation 1.7.9. Hence, with an appropriate range of values of t, and curve can be produced, and standard curve fitting procedures can yield the value for T_1 [63].

Measurement of transverse relaxation T_2 by spin-echo or CPMG

To measure the transverse relaxation rate, one must first generate transverse magnetisation using a standard 90° pulse. The transverse signal begins to relax due to inhomogeneity of chemical shifts, and also due to inhomogeneities in the magnetic field. These magnetic field inhomogeneities increase the rate of loss of transverse signal, such that the observed relaxation rate is faster than the true relaxation rate. However, by application of a 180° pulse after a time $\frac{t}{2}$, all of the spins in the transverse plane are flipped and begin to refocus, cancelling the effect of the inhomogeneous magnetic field. After a time $\frac{t}{2}$ after the refocussing pulse, the spins are completely refocussed and the signal that is detected at this point depends upon the time t according to equation 1.7.10. The short pulse sequence $\frac{t}{2}$-180-$\frac{t}{2}$ is referred to as a spin-echo, or refocussing pulse.

If conformational exchange occurs in the molecule being investigated during the time spine-echo, this can also lead to inaccuracies in the measured T_2. The Carr-Purcell-Meiboom-Gill (CPMG) spin-echo pulse train avoids this problem by repeating the $\frac{t}{2}$-180-$\frac{t}{2}$ pulse multiple times in succession, allowing the effects of

conformational exchange to average out so that the average T_2 of a given nucleus is precisely obtained [63].

Pulsed gradient spin-echo (PGSE) diffusion

To monitor diffusion by NMR, magnetic field gradient pulses are often used. By changing the strength of the magnetic field linearly along the NMR tube (for example, high strength at the top of the tube, and low strength at the bottom), the Larmor frequency of nuclei (in the transverse plane) at different locations in the tube is also altered. This causes a 'corkscrew' effect where nuclei at the top of the sample rotate about the transverse plane more rapidly. Switching the gradient off at this point causes the spins in the sample to remain in this corkscrew. Assuming no translational diffusion, application of an inverted gradient of equal and opposite magnitude for an equal amount of time will directly reverse this corkscrew effect with no overall effect on the resulting signal. The gradient and inverted gradient pulses are known as encoding and decoding pulses, respectively. However, during the delay between the encoding and decoding pulses, translational diffusion is bound to occur, such that nuclei near the top of the sample, which were given a strong increase in magnetic field, will move lower down, and vice versa. This 'scrambles' the corkscrew, and application of the decoding pulse will now yield less signal than would have been observed if no diffusion had taken place. Hence, the rate of diffusion of species in a sample can be measured, by recording multiple spectra and altering the amount of time between the encoding and decoding pulses, as well as the overall strength and duration of the gradient pulses themselves [72].

Heteronuclear SORDID diffusion

Although diffusion is a powerful and crucial tool for the monitoring of sample size, it is relatively insensitive and dependant on the recovery time between scans. The Signal Optimisation with Recovery in Diffusion Delays (SORDID) method is a means of reducing the time taken for a single measurement to be made. In SORDID, diffusion is monitored during the recovery time between two consecutive scans, whereas in standard heteronuclear diffusion experiments, the recovery time occurs after the diffusion monitoring. This effectively halves the amount of time required for a single measurement. This approach is best when trying to measure small diffusion coefficients (of large molecules, for example) of ^{13}C or ^{15}N enriched samples [51].

SOFAST HMQC

The band-Selective Optimised Flip-Angle Short-Transient (SOFAST) HMQC experiment relies on enhanced ^{1}H longitudinal relaxation that is induced by selective excitation only of the spins we wish to observe (amide

protons, for example). This results in many unperturbed protons throughout the duration of the experiment that act as effective sinks for dipolar relaxation, allowing magnetisation to return to equilibrium at a much greater rate, hence increasing the number of experiments that can be collected in a given amount of time. This is extremely useful for large systems, where many scans are needed in order to obtain good quality signal [73].

TROSY NMR

For extremely large systems, such as described in section 1.6.1, conventional decoupled 2-dimensional spectra are inadequate for the collection of high resolution spectra. For a typical, decoupled ^{1}H-^{15}N HSQC/HMQC, one observes a single peak, which is composite of the four peaks in the multiplet that is observed in a non-decoupled spectrum. As a protein increases in size, the larger correlation time τ_c causes these multiplet peaks to have differential linewidths. Some remain relatively narrow, whilst others become extremely broad. In the decoupled spectrum, the broad peaks are added to the narrow peaks, leading to an overall increase in linewidth to the point that the signal is too broad to be observed. However, the narrow components of the multiplet peaks can be selectively observed using phase cycling sequences, eliminating the broad signal entirely. This ultimately leads to a much sharper linewidth by collecting non-decoupled spectra and selectively observing narrow multiplet peaks [74, 75].

1.8.1 The Bloch-McConnell equations and chemical exchange

Describing the rate of change in magnetisation as a function of time in terms of a matrix allows the development of a general method for breaking down and quantifying the various contributions to magnetisation present in a sample.

The Bloch equations [76] describe how magnetisation changes as a function of Larmor frequency and relaxation rate. They can be described in matrix form as given below;

$$\frac{d}{dt}\begin{bmatrix}\hat{E}\\ \hat{M}_z\\ \hat{M}^+\\ \hat{M}^-\end{bmatrix} = \begin{bmatrix} & & & \\ R_1\hat{M}_{z0} & -R_1 & & \\ & & i\omega - R_2 & \\ & & & -i\omega - R_2\end{bmatrix}\begin{bmatrix}\hat{E}\\ \hat{M}_z\\ \hat{M}^+\\ \hat{M}^-\end{bmatrix} \tag{1.8.1}$$

We are interested in the transverse component of two or more species, A and B (and C), undergoing chemical exchange (equation 1.7.1). If we describe the observed transverse magnetisation as ($\hat{M}^+$) as $\hat{M}^+ = \hat{M}_x + i\hat{M}_y$,

then the Bloch equations for free precession of two exchanging species A and B (not accounting for exchange kinetics) can be written thus;

$$\frac{d}{dt}\begin{bmatrix}\hat{M}_A^+\\ \hat{M}_B^+\end{bmatrix} = \begin{bmatrix}-R_{2,A}+i\omega_A & \\ & -R_{2,B}+i\omega_B\end{bmatrix}\begin{bmatrix}\hat{M}_A^+\\ \hat{M}_B^+\end{bmatrix} \tag{1.8.2}$$

We can generate the Bloch-McConnell equations from this by accounting for the kinetics of chemical exchange between A and B, giving;

$$\frac{d}{dt}\begin{bmatrix}\hat{M}_A^+\\ \hat{M}_B^+\end{bmatrix} = \begin{bmatrix}-R_{2,A}+i\omega_A-k_{AB} & k_{BA}\\ k_{AB} & -R_{2,B}+i\omega_B-k_{BA}\end{bmatrix}\begin{bmatrix}\hat{M}_A^+\\ \hat{M}_B^+\end{bmatrix}. \tag{1.8.3}$$

This can be written in more efficient notation as;

$$\frac{d\overrightarrow{\mathbf{M}}}{dt} = \left[i\hat{\boldsymbol{\omega}} - \hat{\mathbf{R}} + \hat{\mathbf{k}}\right]\overrightarrow{\mathbf{M}} \tag{1.8.4}$$

Where $i\hat{\boldsymbol{\omega}}$, $\hat{\mathbf{R}}$ and $\hat{\mathbf{k}}$ are the frequency, relaxation and exchange matrices, respectively, and $\overrightarrow{\mathbf{M}}$ is the magnetisation vector $\left[\begin{smallmatrix}\hat{M}_A^+\\ \hat{M}_B^+\end{smallmatrix}\right]$.

Using this approach allows us to consider various contributions to magnetisation such as transverse relaxation, kinetics and thermodynamics, in a rigorous manner, and this will be useful at later stages in the work presented herein.

Chapter 2

A Novel Approach for the Study of Nascent Chains by NMR Spectroscopy: Design of a New Nascent Chain System

2.1 Fast exchange vs slow exchange

We wish to consider the relative merits of fast exchange against slow exchange in the specific context of studying the co-translational folding of nascent chains by NMR spectroscopy. In order to present such arguments, we first recall the fundamental aspects of how chemical exchange affects the NMR spectra of a protein in a two-state equilibrium between folded and unfolded. We hence consider a precessing nucleus in exchange between two environments, U and F, with forward and reverse reaction rates k_{UF} and k_{FU}.

$$U \underset{k_{FU}}{\overset{k_{UF}}{\rightleftharpoons}} F \tag{2.1.1}$$

By comparing the exchange rate k_{ex} $(= k_{UF} + k_{FU})$ with the difference in larmor frequency $(\Delta\nu)$ of the two states,

$$\Delta\nu = |\nu_U - \nu_F| \tag{2.1.2}$$

we can define the exchange regime into one of three classes, namely slow, intermediate, or fast, as described in equation 1.7.8. These give rise to very different spectra, which must be analysed appropriately according

to the regime they are in. In fast exchange, for a given population distribution of U and F (in molar fractions p_U and p_F), one only observes a single peak, whose observed frequency (ν_{obs}) relative to the frequency of the U and F states (ν_U and ν_F) is a direct measure of p_U and p_F.

$$\nu_{obs} = p_U\nu_U + p_F\nu_F \tag{2.1.3}$$

In slow exchange, it is the relative peak heights that contain the information on the population distribution;

$$p_U = \frac{I_U}{I_U + I_F} \tag{2.1.4}$$

where I_U and I_F are the heights of the resonance peaks corresponding to U and F, respectively. One must be careful with this kind of analysis however, since the intermediate exchange regime lends itself to neither of these two analyses, and inappropriate use of either equation 2.1.3 or 2.1.4 can lead to large errors in population determinations. This then necessitates that the spectra be analysed by solving the Bloch-McConnell equations [77] to fit to the complex lineshapes that can arise [78, 79].

2.2 Chemical exchange on the ribosome

We recall the case of a ^{15}N labelled translation stalled nascent chain in a wholly unfolded conformation, such as is found when the FLN750 domain is stalled 21 amino acids from the PTC (see section 1.6.2). In this case, the extra conformational freedom of the unfolded state leads to a reduced local correlation time relative to the rest of the ribosome, leading to sharper resonances that can be observed by NMR. Since nitrogen has a relatively wide chemical shift dispersion even in the unfolded state, this allows resolution of ^{1}H-^{15}N HSQC crosspeaks, allowing characterisation of the unfolded, ribosome-bound state (U).

On the other hand, stalling of the folded state (F) on the ribosome (such as the dom5 +110 RNC) does not produce observable ^{15}N signal due to greater line-broadening in the conformationally restricted folded state. To observe the folded state, one can utilise specific ^{1}H-^{13}C labelling of isoleucine, leucine, or valine (ILV) methyl group side chains against an NMR-silent ^{2}H-^{12}C background. This gives rise to more intense, sharper spectral lines for large proteins due to the favourable relaxation properties of the methyl group, and utilisation of the methyl TROSY effect [75]. The enhanced sensitivity provided with this labelling strategy has allowed the natively structured protein to be observed on the ribosome [12]. However, methyl side chains have a very narrow chemical shift dispersion for unfolded proteins, and so even though the unstructured state

can also be observed with this method, the severe overlapping of crosspeaks precludes its characterisation.

Consider then a translationally stalled system in exchange between U and F (for example, domain 5 RNCs at linker lengths in the range of 40 amino acids have been seen to adopt this behaviour). In the case of domain 5, which folds on the millisecond timescale, this leads to a slow exchange regime where peak intensity should correspond to the population of U and F. But, since one can only observe either the folded or unfolded state within a single experiment (according to the labelling strategy), one can only determine relative populations using a simplified variant of equation 2.1.4;

$$p_U = \frac{I_{U,obs}}{I_{U,ref}} \tag{2.2.1a}$$

$$p_F = \frac{I_{F,obs}}{I_{F,ref}} \tag{2.2.1b}$$

where $I_{U,ref}$ and $I_{F,ref}$ are intensities obtained from a control spectrum pertaining to the fully unfolded or folded protein. This can be a mixture of ribosomes and domain 5 (or an unfolded domain 5 mutant) at 1:1 molar ratios, or a nascent chain where the stalled protein is known to be 100 % folded or unfolded. Clearly, this necessity for a highly controlled reference sample is problematic, since errors in concentration measurements, and uncertainty regarding nascent chain occupancy and labelling efficiency, leads to large errors in the population measurements. Furthermore, differential relaxation as the nascent chain moves further from the exit tunnel impacts upon linewidth, and hence intensity, further compounding this uncertainty. Finally, nascent chains suffer from inherently weak signal due to the limiting ribosome concentration of 10 μM, and their limited stability precludes signal improvement by simply recording spectra for a long duration of time. Hence, signal to noise ratios are usually weak, such that even if errors can be minimised with respect to knowledge of sample concentration, there is still an inherent uncertainty originating in the weakness of the signal itself.

Consider instead a system in fast exchange on the chemical shift timescale. This regime benefits greatly from the fact that the population distribution now alters peak position (equation 2.1.3) as opposed to peak height (equation 2.1.4). Measurements of frequency benefit by having less dependence upon the signal level in a sample. Hence, even at low signal to noise, frequency determination remains accurate, eliminating all errors associated with concentration and occupancy measurements. Differential relaxation as a function of chain length also has little effect on error, since relaxation should not affect peak position.

In order to verify the hypothesis that fast exchanging systems are beneficial for the study of nascent chains by NMR, we describe in the next section a series of lineshape simulations of differing exchange regimes for

a system in chemical exchange between two states.

2.3 Evaluation of Slow and Fast Exchange Regimes via Lineshape Simulations

Simulations were carried out to replicate a precessing nucleus in exchange between two states U and F as described in equation 2.1.1. The simulations were carried out in matlab using a specifically designed script. The full script used for the simulation and plotting of data is provided in Appendix A.

The change in magnetisation over time for our system can be described as the time derivative of the following Bloch-McConnell matrix;

$$\frac{d}{dt}\begin{bmatrix} M_U^+(t) \\ M_F^+(t) \end{bmatrix} = \begin{bmatrix} -(R_{2U} - i\omega_U) - k_{ex}p_F & k_{ex}(1-p_F) \\ k_{ex}p_F & -(R_{2F} - i\omega_F) + k_{ex}(p_F - 1) \end{bmatrix} \cdot \begin{bmatrix} M_U^+(t) \\ M_F^+(t) \end{bmatrix} \tag{2.3.1}$$

which can be expressed in a simplified form as

$$\frac{d\vec{\mathbf{M}}(t)}{dt} = \mathbf{A} \cdot \vec{\mathbf{M}}(t) \tag{2.3.2}$$

the solution to which is the matrix exponential.

$$\vec{\mathbf{M}}(t) = \mathrm{e}^{\mathbf{A}t}\vec{\mathbf{M}}(0) \tag{2.3.3}$$

To generate a spectrum, the expression in equation 2.3.3 was put into matlab as a matrix, with chosen values for each variable (R_2, k_{ex}, etc), and the value of the matrix was numerically evaluated for a range values t from 0 to 205 ms in increments of 0.2 ms. This generated a simulated free induction decay (FID), which was then treated as a real, time domain NMR signal. To simulate realistic experimental conditions, the signal was modified by inclusion of additive white gaussian noise such that the FID closely resembled real data. This was performed using the 'awgn' Matlab function. White gaussian noise was chosen as it represents the noise that always arises from stochastic processes in a sample, such as thermal noise. Although other sources of noise could also have been considered, such as phase noise, we chose in this case to investigate the simplest case of stochastic noise.

After the addition of noise, an apodisation window function (*apod*) was applied,

$$apod = \cos^2\left(\frac{\max(t)\pi}{2t}\right) \tag{2.3.4}$$

since this was found to yield slightly sharper linewidths. Application of a complex Fourier transform (Matlab function fft) then yielded the frequency domain NMR signal. The apodisation window function is an optional addition to the simulation.

Simulation parameters

Parameters such as chemical shifts and transverse relaxation were carefully selected to reflect the behaviour of the two main types of nucleus that are commonly detected in RNC NMR studies. Namely, backbone amide protons and methyl side chain protons of ILV residues (note that methyl protons are typically measured against a deuterated background).

Two sets of relaxation rates were chosen to correspond to these scenarios. Methyl protons have been observed to have relaxation rates of approximately 50 s^{-1} in the folded, ribosome bound state, while in the unstructured form this is difficult to measure due to significant peak overlap, although it can be estimated to be 25 s^{-1} based on observations of amide protons. In the unfolded state, an R_2 for amide protons of 25 s^{-1} is typically observed [11, 14], though this tends to vary. Since the structured state on the ribosome cannot be observed via amide protons, this was estimated to be 300 s^{-1}, as this is the minimum value required to generate extremely broad signal. However, since this signal cannot be observed, it could be much greater than 300 s^{-1}.

For each nucleus, ν_F was set to 350 Hz and ν_U was 0 Hz. At 700 MHz, this would represent a $\Delta\delta_{F-U}$ of -0.5 ppm. Spectra were then generated as described above, using two different exchange regimes, $k_{ex} = 10$ and 10^5 s^{-1}. For each exchange regime, p_F was incremented from 0 to 1 in increments of 0.1.

Hence, if we consider a single experiment as a titration where spectra are recorded for different fractional populations of U and F (by addition of a chaotrope, for example), then a total of 4 experiments were simulated, namely methyl and amide proton spectra for systems in slow and fast exchange. These simulations were then repeated at 8 different noise levels, bringing the total number of experiments to 32. The simulation results are summarised in Figures 2.1 to 2.4.

To determine the maximum signal to noise ratio per experiment, the maximum height of the sharp peak (peak U) at a population of 100 % was measured relative to the standard deviation of noise in the spectrum,

leading to approximate signal to noise ratios of 4, 7, 13, 21, 46, 71, 130 and 245 for the titrations. The true value of these ratios varies for different values of p_F, as peak heights change. The noise levels as given are described in the remainder of this section as noise level A, B, C, D, E, F, G and H, respectively, reflecting the representation in Figures 2.1 to 2.4.

The spectra were then treated to a simple population analysis according to whether they were in fast or slow exchange, using equations 2.1.3 and 2.2.1, respectively. In slow exchange, the intensity of the unfolded peak was measured as a function of p_F for the amide proton simulations, and the folded peak intensity was measured in the case of methyl protons, reflecting the fact that the other peaks cannot be observed experimentally. Uncertainty in peak height was calculated using the standard deviation of the noise level in each spectrum to determine the relative error.

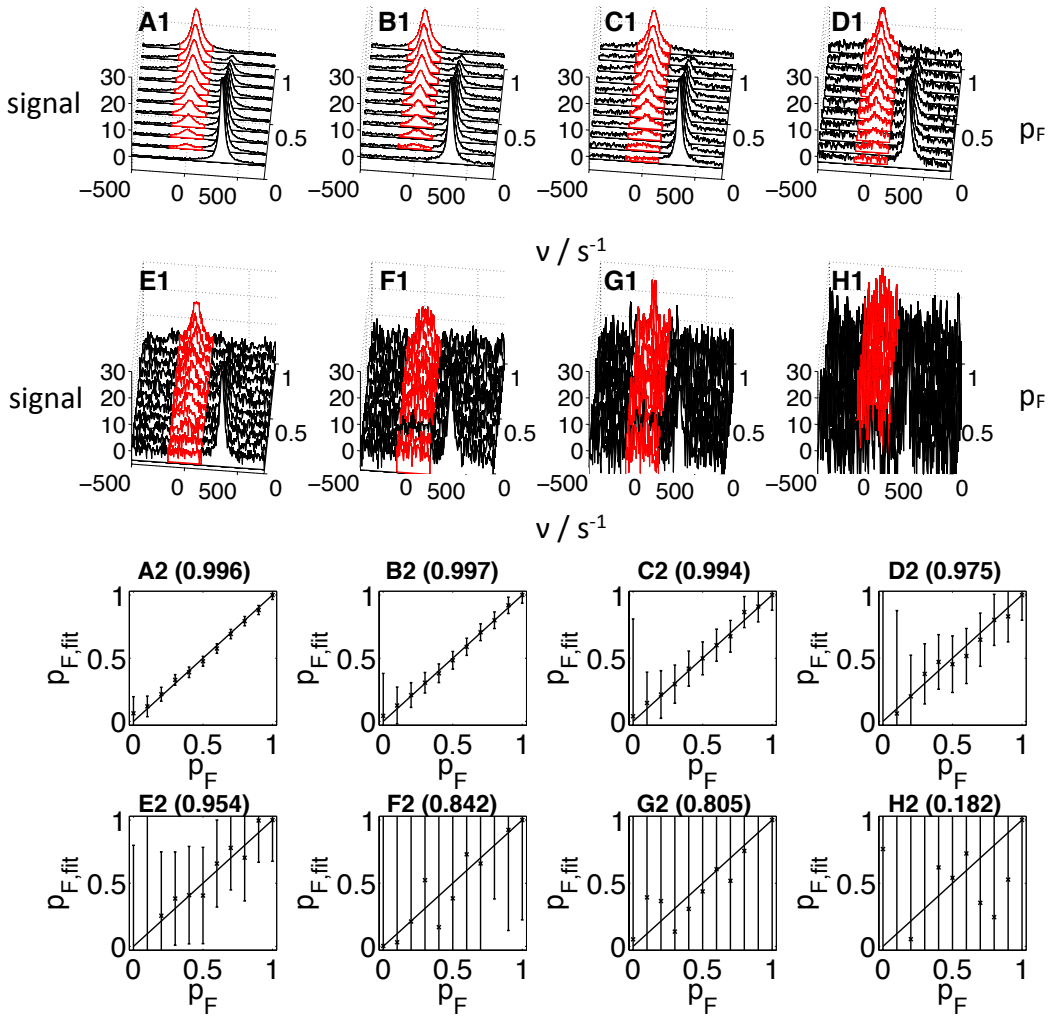

Figure 2.1: *Simulated spectral lines for methyl protons in slow exchange, and extraction of p_F values with associated error.* **A1-H1** Simulated spectra, with incremental amounts of noise from A to H. **A1** and **A2** represent the spectra (**A1**) and corresponding plot (**A2**) of true vs observed values for the fraction of F, p_F for the highest signal to noise ratio, while **H1** and **H2** represent the lowest signal to noise ratio. The numbers shown in brackets above **A2** to **H2** represent the coefficient of determination, R^2 between the observed and true value for p_F for the corresponding plot. Taking this value along with the observed error in measurement of p_F values allows for an evaluation of the relative quality of each individual fit. Populations were determined using equation 2.2.1.

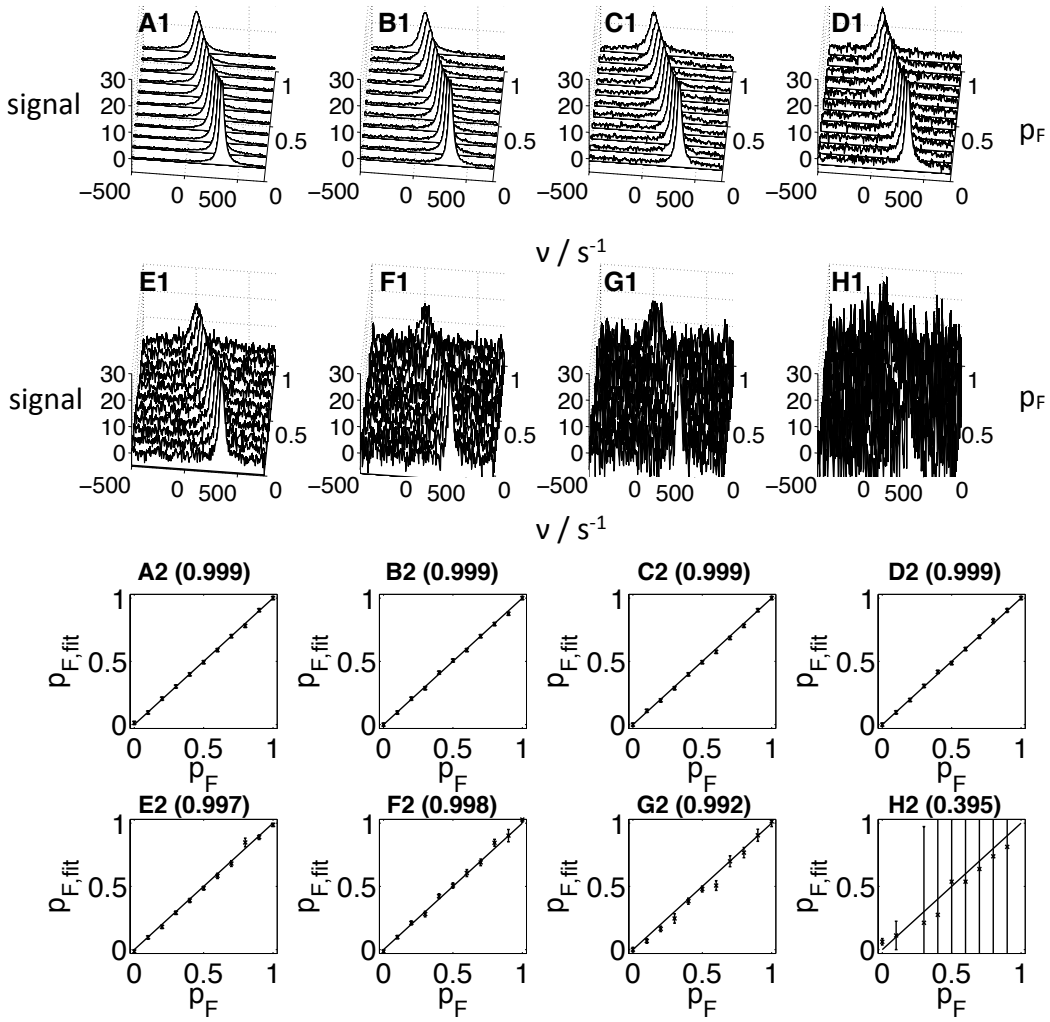

Figure 2.2: *Simulated spectral lines for methyl protons in fast exchange, and extraction of p_F values with associated error.* **A1-H1** Simulated spectra, with incremental amounts of noise from A to H. **A1** and **A2** represent the spectra (**A1**) and corresponding plot (**A2**) of true vs observed values for the fraction of F, p_F for the highest signal to noise ratio, while **H1** and **H2** represent the lowest signal to noise ratio. The numbers shown in brackets above **A2** to **H2** represent the coefficient of determination, R^2 between the observed and true value for p_F for the corresponding plot. Taking this value along with the observed error in measurement of p_F values allows for an evaluation of the relative quality of each individual fit. Populations were determined using equation 2.1.3.

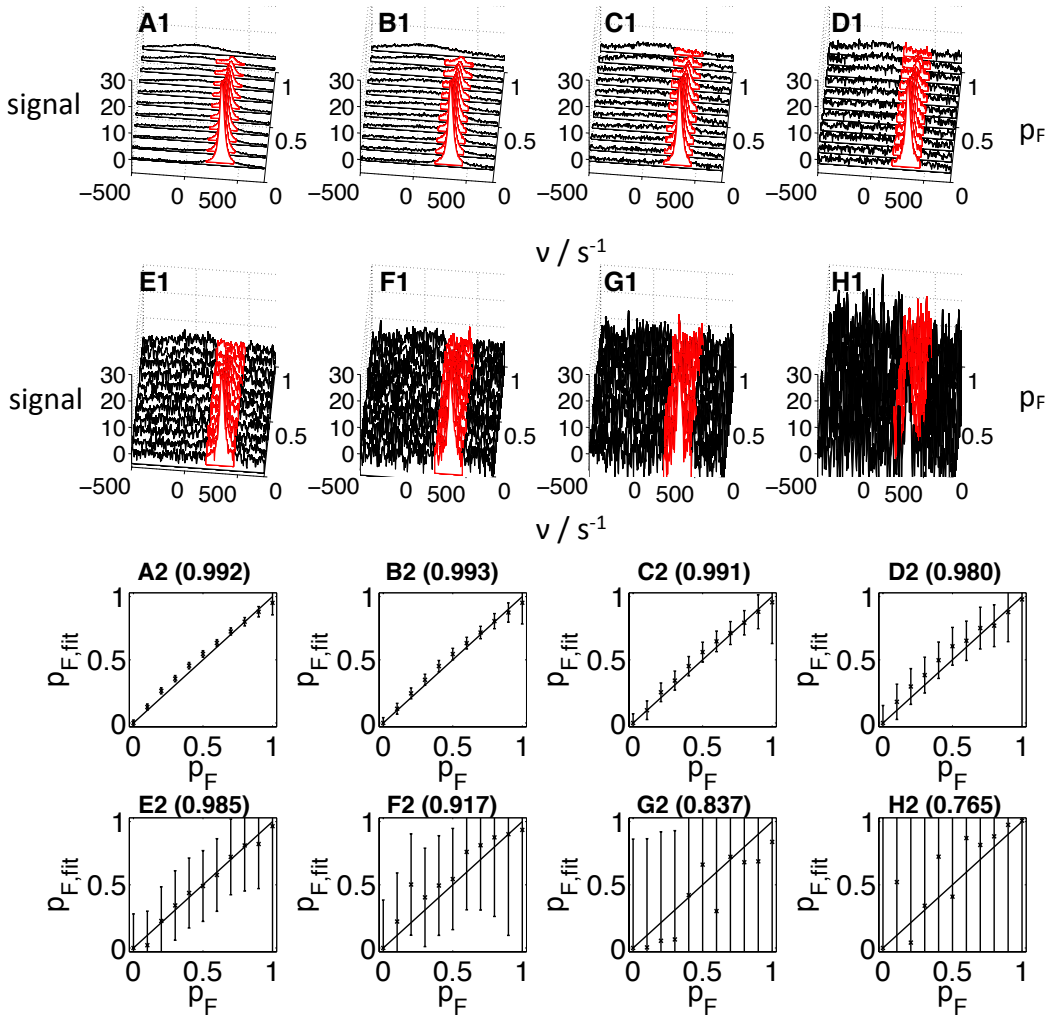

Figure 2.3: *Simulated spectral lines for amide protons in slow exchange, and extraction of p_F values with associated error.* **A1-H1** Simulated spectra, with incremental amounts of noise from A to H. **A1** and **A2** represent the spectra (**A1**) and corresponding plot (**A2**) of true vs observed values for the fraction of F, p_F for the highest signal to noise ratio, while **H1** and **H2** represent the lowest signal to noise ratio. The numbers shown in brackets above **A2** to **H2** represent the coefficient of determination, R^2 between the observed and true value for p_F for the corresponding plot. Taking this value along with the observed error in measurement of p_F values allows for an evaluation of the relative quality of each individual fit. Populations were determined using equation 2.2.1.

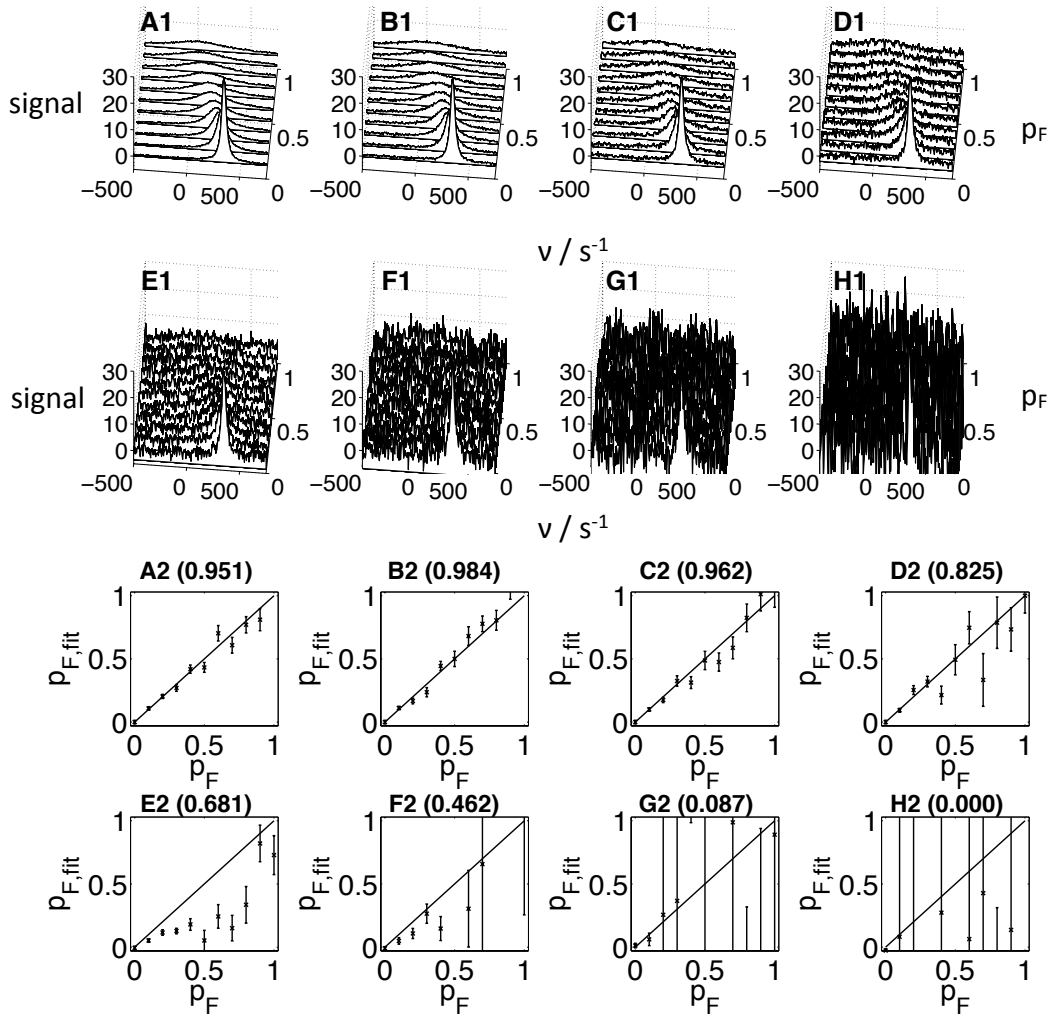

Figure 2.4: *Simulated spectral lines for amide protons in fast exchange, and extraction of p_F values with associated error.* **A1-H1** Simulated spectra, with incremental amounts of noise from A to H. **A1** and **A2** represent the spectra (**A1**) and corresponding plot (**A2**) of true vs observed values for the fraction of F, p_F for the highest signal to noise ratio, while **H1** and **H2** represent the lowest signal to noise ratio. The numbers shown in brackets above **A2** to **H2** represent the coefficient of determination, R^2 between the observed and true value for p_F for the corresponding plot. Taking this value along with the observed error in measurement of p_F values allows for an evaluation of the relative quality of each individual fit. Populations were determined using equation 2.1.3.

In the case of fast exchange, the frequency of the maximum point in the spectrum was recorded at all values of p_F. To determine uncertainty, the simulations were repeated twenty times, and the maximum value was recorded in each case. The introduction of random noise altered the observed peak maximum for each repetition, and the standard deviation of this gave the uncertainty in the measured frequency of the peak (see section 7.12 for error propagation techniques). R^2 correlation coefficients were also determined

The results of this analysis at the various noise levels is summarised in Figures 2.1, 2.2, 2.3 and 2.4.

Combining the R^2 values with the error bars on each plot allows comparisons to be made regarding the accuracy and precision between the fast and slow exchange regimes for the two types of nuclei. Consider the methyl proton simulations (Figures 2.1 and 2.2). It is clear that for fast exchange, the R^2 value remains high even at very low signal to noise values, whilst for slow exchange the coefficient begins to steadily drop in value (compare plots C2 to G2 in both sets of figures). It is also clear that the error in p_F becomes restrictively high in the case of slow exchange whilst the same noise level in fast exchange has little impact upon the error (consider plots E2 and F2 of Figures 2.1 and 2.2).

Alternatively, the amide proton simulations show a different pattern of behaviour. Considering plots **D2** of Figures 2.3 and 2.4, we see that in slow exchange there is a strong R^2 value of 0.980 and an error in the measured p_F value that is small enough for a clear shift in population to be observed, whilst in fast exchange, the R^2 value is notably weaker (0.825), and the error bars are unreliable. Plots E2 through to H2 then show a rapid decline in quality of fit for both regimes, with very large error bars in the case of slow exchange (poor quality precision), and low R^2 values in the case of fast exchange (poor quality accuracy).

The inaccuracies observed for the amide proton fast exchange spectra can be attributed to the severe broadening observed in the folded state. This makes measurement of the peak position very difficult as the signal is so close to the baseline. Therefore, although it seems clear that the fast exchange regime provides a clear advantage in the case of relatively slow relaxing species such as methyl protons against a deuterated background, when relaxation rates become too high, this advantage is less clear. Nevertheless, the superiority observed in the case of the methyl proton simulations acts as an important initial proof of principle that fast exchanging systems can provide detailed information on populations at low signal to noise.

It is important to also consider how these plots represent what is observed for real nascent chains. Consideration of in-lab data acquired over the years for many systems suggests that a high quality nascent chain sample with appropriate controls would have signal to noise values that appear similar to what is observed in plots D1-2 and E1-2 of Figures 2.1 to 2.4. Clearly, in the case of methyl protons, fast exchange displays a huge advantage over slow exchange, but for amide protons this is perhaps not so clearly the case.

These simulations do not take into account the possibility that relaxation can change as a function of chain length for both the folded and unfolded states, thus introducing a separate component that affects the intensity of a resonance. Also not considered were uncertainties regarding concentrations and occupancy measurements, which would further affect intensities, but not chemical shifts.

Hence, the major complementary argument to consider alongside the simulations is that experimental conditions introduce inherent uncertainties into a sample, such as concentration and occupancy, which correspondingly leads to large uncertainties in peak intensity. Peak positions, however, are not affected by these factors, and so properties of the system that impact upon observed chemical shifts can be more reliably quantified, even in the case where concentration or occupancy is unknown. It must be noted however, that this superiority is dependent to a large extent on the ability to record ν_U and ν_F to high accuracy in isolated systems. This assumption is reasonable provided that the folded and unfolded states can be observed in solution under conditions as similar as possible to those of the nascent chain.

2.4 Chapter summary

To summarise, we have shown how consideration of the physical properties of a particular peptide, namely the folding exchange rate, is an important step in appropriate choice of a nascent chain system. Some fast folding proteins have been studied co-translationally in the literature, such as in the work by Marino et al. [42]. However, the fast exchange regime has not been exploited for these peptides by NMR, and therefore we wish to explore the potential this approach has for providing detailed structural information about folding on the ribosome. We now discuss approaches towards identifying the fast folding protein candidates that will be used to achieve this goal.

Chapter 3

Identification of Fast Folding Proteins and Subsequent Folding Characterisation

Due to the considerations of fast exchange against slow exchange, it was decided to design and develop a complementary system to the FLN750 RNC, comprising a protein that folds extremely fast on its chemical shift timescale. To this end, a literature survey was carried out to determine fast-folding candidates. Some of the proteins identified are shown in Table 3.1. Bearing in mind that the ribosome can reduce folding rates of some nascent chains (such as T4 lysozyme [2]), proteins with 'ultrafast' folding rates (approaching the microsecond timescale) were preferred.

To evaluate whether a given exchange rate would yield the fast exchange phenomenon, the literature determined folding rates were compared against a $\Delta\delta$ of 1 ppm at 700 MHz. This is an estimate of the typical chemical shift difference between folded and unfolded states for an amide proton. To qualify as fast exchange with minimal chances of entering the intermediate exchange regime, it was decided that the folding rate should be at least two orders of magnitude greater than the frequency difference between the folded and unfolded states, $\Delta\omega$. A number of the options in table 3.1 have folding rates that are two orders of magnitude greater than $\Delta\omega$, making them particularly promising.

Of the proteins identified in Table 3.1, two were subjected to more detailed investigation. Namely, the HP35 villin headpiece and the GA module. They are characterised by having very rapid folding times, and also have established characterisations of their folding kinetics and thermodynamics in the literature. Although there were numerous candidates to choose from, these peptides represented promising kinetics as

Protein	PDB code	Length	Size / kDa	τ_f / µs	Temp / °C	$\frac{k_f}{\Delta\omega}$ [a]	Reference
Engrailed Homeodomain	1ENH	54	6.6	26	25	9	[80]
Trp-cage	1L2Y	20	2.2	4.0	22.7	57	[69]
Trp^2-cage	$\sim$	20	2.3	1.0	22.7	227	[81]
GA module	1PRB	47	5.9	1.0	94.5	227	[82]
α_3D	2A3D	73	8.1	3.2	50	71	[83]
HP-35 Villin Headpiece domain	1YRF	35	4.1	0.7	27	325	[84]
B-domain of protein A	1BDD	60	6.8	10	-	23	[85] [b]
WW domain	1I6C	24	3.0	78	41.7	2.9	[86]

[a] Calculated at 700 MHz and assuming a $\Delta\delta$ of 1 ppm (i.e., $\Delta\nu = 700$ Hz, $\Delta\omega = 4398$ s^{-1}).
[b] Data given is for the Y15W mutant.

Table 3.1: *List of proteins with fast folding kinetics.* The folding time τ_f is given by the reciprocal of the folding rate, k_f, and is true for the temperatures provided in the table.

well as a reasonable size. Too small, and one faces issues with purification of the isolated proteins for control experiments. Too large, and this can restrict computational studies which can provide additional information to complement experimental work. It was also found that most fast folding proteins are relatively small, hence the lack of larger species.

The DNA plasmid for the HP36 was obtained as a donation from Professor Dan Raleigh, and GA module DNA was donated by Dr Sarah Dodd (née Lejon).

3.1 The GA module

The protein PAB is a human serum albumin (HSA) binding protein identified in HSA-binding strains of *Finegoldia magna* (formerly *Peptostreptococcus magnus* [87]). The GA module is a 45 residue albumin binding domain within PAB that has 60 % homology with protein G, and was hence termed the protein **G**-related **A**lbumin binding domain, or GA.

The structure has been well established as a three helix bundle by crystallography [88] and NMR [89, 90], and characterisation of its interaction with albumin has also been carried out [91, 92]. Of particular interest are the studies of its folding behaviour. Infra-red temperature jump (IR T-jump) studies of the wild type protein have shown it to have a folding time of $\sim$6 µs. This folding time was measured using a temperature jump between 84.3 and 94.5 °C[82]. The same method has also been used to show that a double mutant of the GA module, (K5I/K39V) has a folding time constant τ_f of $\sim$1 $(\mu s)^{-1}$ [82]. This double mutant was designed based on simulation work carried out by Zhu et al. [93] with the intention of fine-tuning the folding rate by introducing mutations that alter the hydrophobic core. Other simulations by Takada [94] and Lei

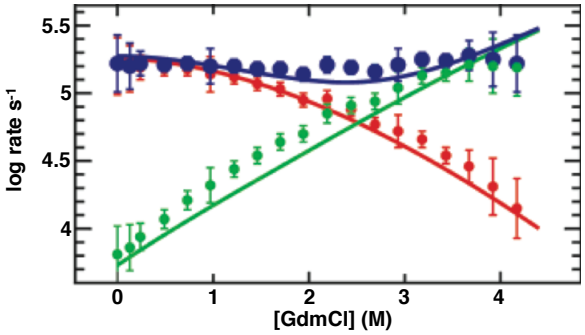

Figure 3.1: *Chevron plot of folding kinetics of HP36 headpiece as a function of guanidine chloride concentration.* The folding rate, k_{uf}, shown in red, gradually decreases with increasing [GdmCl] while the unfolding rate k_{fu} (green) increases. The overall exchange rate, k_{ex} (shown in blue, $= k_{fu} + k_{uf}$), appears to change by a small degree, but overall remains in a similar range throughout the titration. Figure was extracted directly (with small aesthetic alterations) from Cellmer et al. [96].

at el. [95] have also confirmed, from a computational standpoint, that the GA module has extremely fast folding kinetics.

It must be borne in mind, however, that these folding studies were carried out using T-jump experiments, and the elevated temperatures at which they were performed could lead to increased rates than would be observed at 25 °C. This is an important caveat, since isolated ribosome complexes are not stable at temperatures significantly greater than 25 °C. The high temperatures in the literature were used because the GA module appears to have very high thermal stability. Johansson et al. [89] attempted to determine the melting point, T_m of the domain, but were unable to find a precise temperature for this value, although they did determine that the module appears to begin unfolding above 70 °C.

Such high folding temperatures are often indicative of thermophilic proteins. However, the *Finegoldia magna* strain that the GA is derived from is not thermophilic, and so it appears in this case that the GA is simply a very stable small domain of a larger protein.

3.2 The villin headpiece, or HP36 domain

The HP36 is a 36 residue domain that comprises the C-terminus of the villin headpiece, and is the smallest naturally occurring sequence known to fold cooperatively. The folding kinetics of the HP36 have been studied extensively by a range of techniques, including IR T-jump [97], laser induced fluorescence [84, 98] and NMR line-shape analysis [99]. All such studies have confirmed that this domain folds on a microsecond timescale at room temperature conditions. Due to its extremely rapid kinetics and small size, it has become

a very popular domain for computational studies [100–102], which complement a large and growing body of experimental work on its folding properties [96, 103–107].

What makes the HP36 a very promising candidate, besides its ultrafast folding rate, is the extent to which its folding rate does not appear to alter significantly with denaturant concentration, or mutational insertions designed to destabilise the peptide. Figure 3.1 shows the chevron plot of the HP36, obtained from Cellmer et al. [96]. This chevron plot shows the measured folding and unfolding kinetics of the HP36 domain at a range of concentrations of guanidinium chloride. Typically, as the concentration of denaturant reaches levels that cause equal populations of the U and F states, the overall folding exchange rate, k_{ex}, is reduced relative to the fully folded or unfolded proteins. While that is also the case here, it should be noted that the minimum exchange rate is still in the microsecond regime, meaning that we are likely to still be able to observe fast exchange in an NMR spectrum. One can suppose that this 'stability' in the folding rate rate of HP36 gives it a distinct advantage over other proteins where the kinetics of folding as both the folded and unfolded states become populated is greatly reduced. Alongside this, folding studies of the F47L/F51L double mutant, which is significantly destabilised relative to the wild type, indicate that it has similar overall kinetics [97]. Again, this makes the HP36 a very promising candidate, as it is likely to remain in fast exchange over a range of conditions.

Although extremely useful for studies of folding, care must be taken in the production of this peptide since its small size renders it susceptible to proteolysis. Although it is possible to synthesise such a small amino acid stretch, this is still expensive and not suitable for isotopic labelling or the investigation of ribosome-nascent chain complexes. Hence, for the purpose of NMR studies, Bi et al. [108] developed an efficient method for production of the HP36 by producing a construct with an N-terminal tag, cleavable by factor Xa, consisting of the N-terminus of the bacterial ribosome protein L9 (NTL9);

NTL9-Xa-HP36

The NTL9 is very stable, and is easily purified by ion exchange chromatography, allowing it to be used to protect the HP36 during purification before cleavage with factor Xa to produce the villin headpiece with good yield, after further purification to remove the NTL9 and factor Xa.

3.3 Characterisation of isolated proteins

3.3.1 Structure and folding characterisation by circular dichroism

Initial secondary structure investigations were carried out using far-UV circular dichroism (CD). Each protein was dissolved to 10 μM in a buffer at pH 7.5 (GA module in sodium acetate, HP36 in Tris), and a CD spectrum acquired in the wavelength range of 200-250 nm. Spectra were acquired in steps of 1 nm, with 2 seconds per point. CD spectra of both proteins appear consistent with α-helical structures. In particular, the minimum at 222 nm is indicative of this secondary structure motif.

It is important to establish the conditions under which unfolding can occur and to ensure that the unfolded state can be populated to completion, since this is a required end point for accurate population calculations by NMR. Folding characterisation was hence carried out using temperature melts in order to determine the conditions under which the proteins were stable. This was carried out by executing CD temperature melts over a range of concentrations of urea.

Multiple samples were prepared for each peptide, with increasing concentrations of urea. Initial CD measurements on samples with no urea showed a minimum at 222 nm characteristic of α-helical content. Each sample was then subjected to a temperature melt, with a CD spectrum being acquired for each degree increment in temperature (1 nm steps, 0.5 seconds per step) (see Figure 3.2). For both the GA module and the HP36, the CD value at 222 nm was used to monitor the extent of unfolding since both peptides have secondary structures consisting entirely of helices and turns (Figure 3.2).

Although complete unfolding was achieved for both peptides, the GA module proved far more difficult to unfold than the HP36. Whilst the HP36 was almost entirely unfolded in 0 M urea at 90 °C, the GA required between 6 and 10 M urea alongside temperatures above 80 °C in order to obtain a saturated CD signal.

3.3.2 NMR urea titrations

After the initial estimation of the folding ranges of the two domains, the urea titrations were repeated using NMR spectroscopy in order to evaluate folding kinetics and to determine the exchange regime of the peptides as unfolding occurred. Titrations were carried out for the GA module and HP36 domain. All spectra were acquired in 100 mM NaOAc at pH 5.0. The pH was lowered to increase the number of resolvable peaks by preventing amide proton exchange broadening that can occur under more basic conditions. The progression of folding was followed by 1D ^{1}H spectra (Figure 3.3) and ^{1}H-^{15}N HSQC spectra (Figures 3.4, 3.5 and 3.6).

Since the GA module required relatively high temperatures to be unfolded, the urea titration was carried

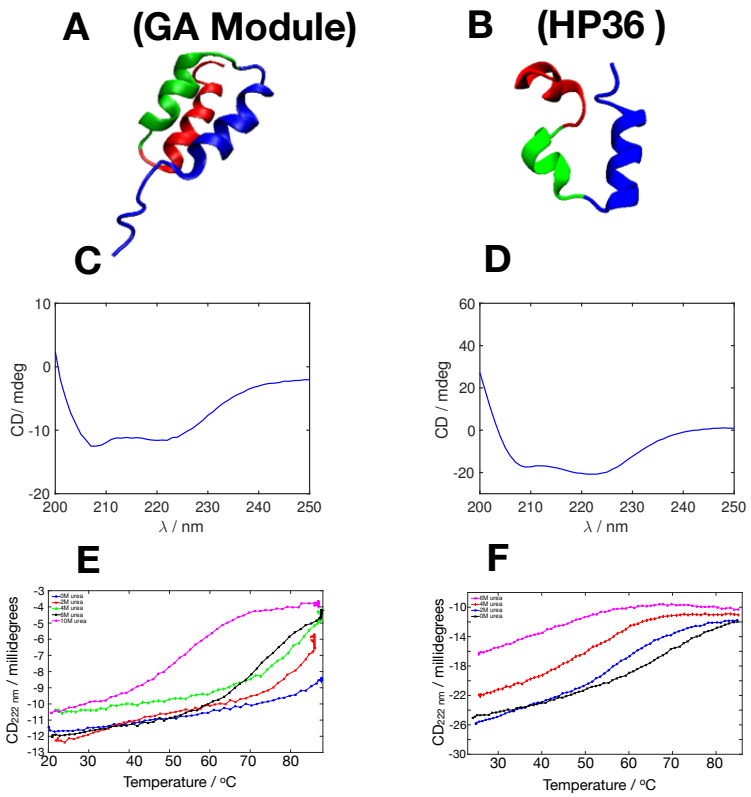

Figure 3.2: *Circular dichroism spectra and urea temperature melts of the GA module and the HP36-NTL9 fusion protein.* **A-B:** Tertiary structures of the GA module (**A**) and HP36 (**B**) domains. Structures were rendered using the visual molecular dynamics (VMD) software [109]. **C-D:** CD spectrum of the GA module (**C**) and the HP36-NTL9 protein (**D**) **E-F:** CD temperature melts of the GA (**E**) and HP36-NTL9 (**F**), at incremental urea concentrations. The CD value at 222 nm was used to monitor unfolding throughout the temperature ramp. Plots were produced using matlab.

out at 55 °C, the highest safe temperature allowed for the NMR spectrometer. According to the CD melts, this temperature and a urea concentration of 10 M corresponds to at least a partially unfolded form of the protein. Spectra were acquired at 800 MHz.

Since the HP36 could be unfolded under milder conditions, this titration was carried out at 25 °C (at 500 MHz), and was then repeated at 55 °C to ensure that complete unfolding was attained. The titration was in fact carried out on the NTL9-HP36 fusion protein, as the folding kinetics of either domain were expected to act independently.

1D spectra

Figure 3.3 shows the high field methyl region of the GA module at increasing concentrations of urea. As the urea concentration begins to increase, the high field shifted peak becomes significantly broadened until it can no longer be observed.

For the HP36 at 328K, (Figure 3.3, top), although some broadening of the highlighted methyl proton is observed, it is largely preserved throughout the titration. The chemical shift of this peak changes in a clearly sigmoidal fashion, which is also indicative of a fast exchange process. However, the same peak at 298 K shows behaviour more like fast-intermediate exchange, as the peak shifts initially, and then broadens out almost completely at 5M urea (Figure 3.3, bottom left). This behaviour is also observed for the GA module, even at 55 °C (Figure 3.3 bottom right).

2D spectra of the GA module

2D spectra of the GA module are also indicative of fast-intermediate or slow exchange. Although some peaks appear to shift gradually, this is likely due to the addition of urea. The broadening of folded cross-peaks is simultaneous with the appearance of peaks corresponding to the unfolded domain (Figure 3.4). In particular, the G22 peak was seen to broaden out and reappear again in a clean region of the spectrum (Figure 3.4). This is consistent with slow exchange, suggesting this protein may be unsuitable for further investigation as a fast folding nascent chain.

Inspecting the chemical shift change of the G22 cross peak, we can see that the overall chemical shift change, $\Delta\delta$, is approximately 0.37 ppm at 800 MHz in the ^{1}H dimension, corresponding to a frequency difference $\Delta\omega$ of $1860\,\mathrm{s}^{-1}$ ($0.37 \times 800\,\mathrm{MHz} \times 2\pi$). In the ^{15}N dimension, the difference of 1.8 ppm corresponds to a frequency difference of $900\ \mathrm{s}^{-1}$ ($1.8 \times \frac{800\ \mathrm{MHz}}{10} \times 2\pi$), since the gyromagnetic ratio of $^{15}\mathrm{N}(\gamma_{^{15}\mathrm{N}})$ is one tenth that of $\gamma_{^{1}\mathrm{H}}$. If we suppose a slow exchange regime, this suggests the folding exchange rate of the GA under these conditions is much less than $900\ \mathrm{s}^{-1}$, or that the folding time τ_f is greater than 1.1 ms.

2D spectra of HP36

The HP36 2D spectra at 328K show some evidence of fast exchange across the urea titration (See Figures 3.5 and 3.6). However, at 298K, there is evidence of a fast-intermediate exchange regime, where in the 1D spectra (Figure 3.3 and 3.6), the high field methyl peak that was in fast exchange at 328K is clearly broadening out at mid-range urea concentrations.

By inspection of the spectra of the HP36 titrations in the same manner as described above for the GA

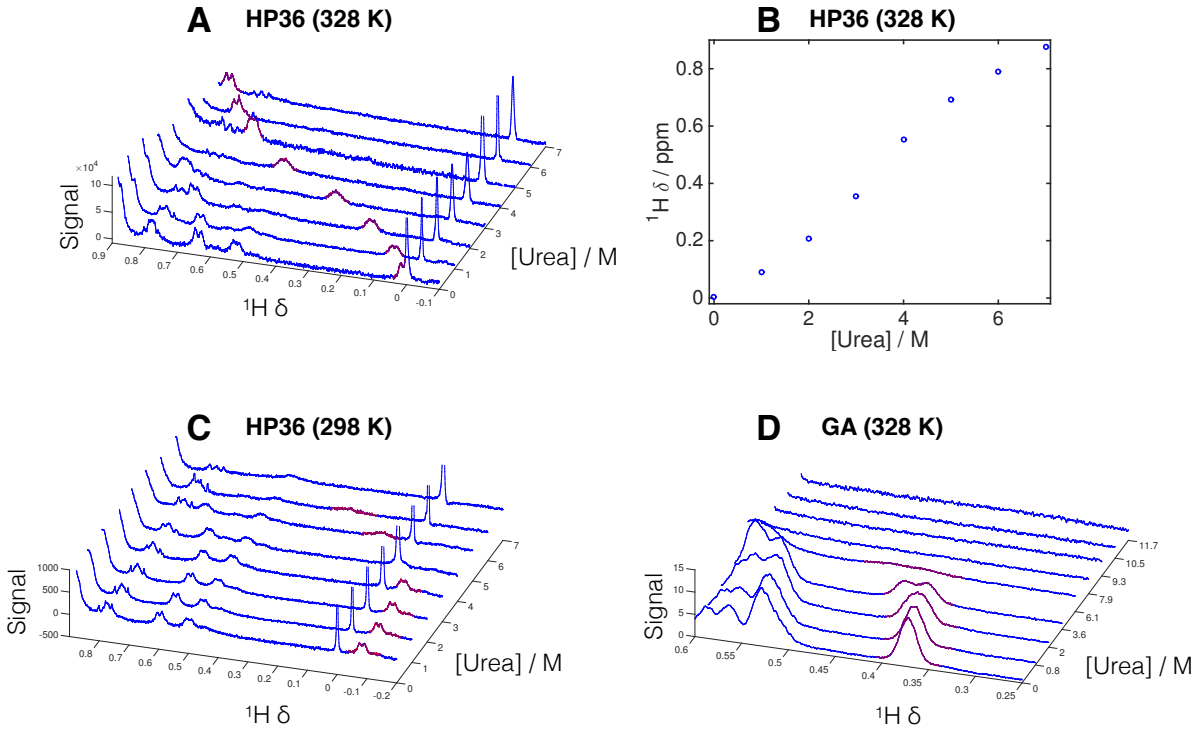

Figure 3.3: *1D spectra of the GA and HP36 as a function of urea concentration.* **A:** Change in high field methyl peak in HP36 with urea concentration, and **B:** plot of change in chemical shift of the highlighted peak. Spectra were acquired at 328K and 500 MHz. **C:** Change in high field methyl peak in HP36 with urea concentration at 298K. Spectra were acquired at 500 MHz. **D:** Change in high field methyl peak in GA module with urea concentration. Acquired at 328K and 800 MHz. Plots were produced using the matlab software [67].

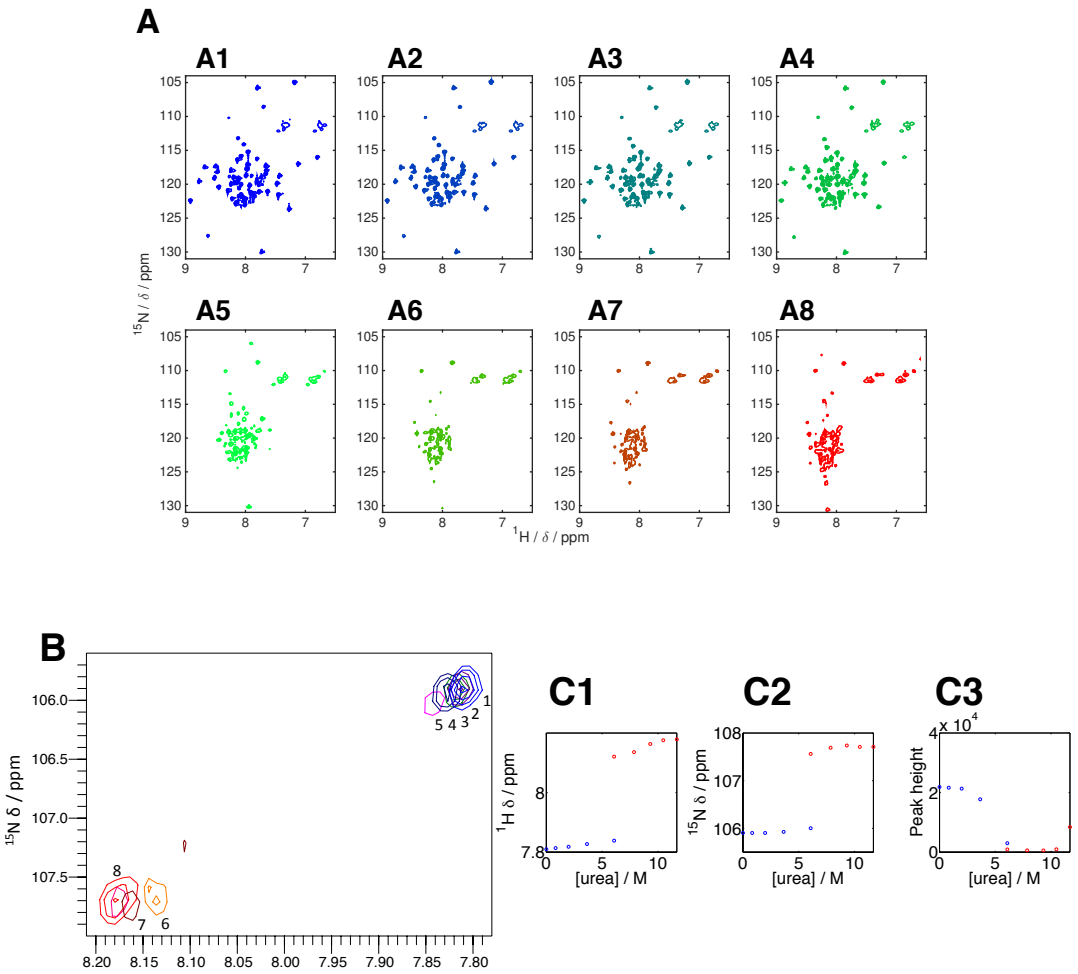

Figure 3.4: *2D spectra and analysis of the urea titration of the GA module.* **A:** Amide region of the ^{15}N HSQC of the GA module at incrementing concentrations of urea. Urea concentrations are as follows; **A1:** 0 M, **A2:** 0.8 M, **A3:** 2.0 M, **A4:** 3.6 M, **A5:** 6.1 M, **A6:** 7.9 M, **A7:** 9.3 M, and **A8:** 10.5 M. Note how the transition from dispersed to non-dispersed is sharp, indicating slow exchange. Spectra were acquired at 800 MHz and 328 K. **B:** ^{1}H-^{15}N HSQC spectra of the G22 crosspeak for each urea concentration. Chemical shifts and peak heights were determined from the spectra and are plotted on the right (Figures **C1**, **C2** and **C3**). The red circles correspond to the appearance of a new peak, with no apparent intermediate chemical shifts.

module, we can see that at 328 K, the A57 peak frequency differences are 310 Hz in the ^{1}H dimension, and 35 Hz in the ^{15}N dimension. At this temperature, we observe fast exchange, indicating that the exchange rate is greater than 1900 s^{-1} (τ_f < 0.5 ms). At 298 K, the same analysis reveals a $\Delta\nu$ of 2500 Hz in the ^{1}H dimension, and 600 Hz in the ^{15}N dimension. In this case, assuming slow exchange indicates that $k_f < 600$ s^{-1}, and $\tau_f > 1.67$ ms.

3.4 Urea titration discussion

If we assume that a simple two-state exchange model (equation 2.1.1) can be applied to the observations made in each titration, then there appears to be a remarkable discrepancy between the kinetics of folding observed here and the microsecond timescales observed in the literature. However, there is a possibility that other events besides those described by equation 2.1.1 are occurring. Consider Figures 3.3C and 3.3D, showing the 1D spectra of the urea titration of the HP36 at 298 K and the GA module at 328 K, respectively. In both plots, as the urea concentration increases above 5 M, substantial line-broadening occurs. Because there are no observable peaks in this range, the origin of this line-broadening cannot be evaluated, and may be due to slow or intermediate exchange. Alternatively, there may be a third state that is populated on the folding pathway (equation 1.2.2), which can lead to complex exchange kinetics, and substantial NMR line-broadening. Another possibility is the formation of a molten globule state. A molten globule is a compact form of a protein with strong secondary structure elements and little to no tertiary structure. The formation

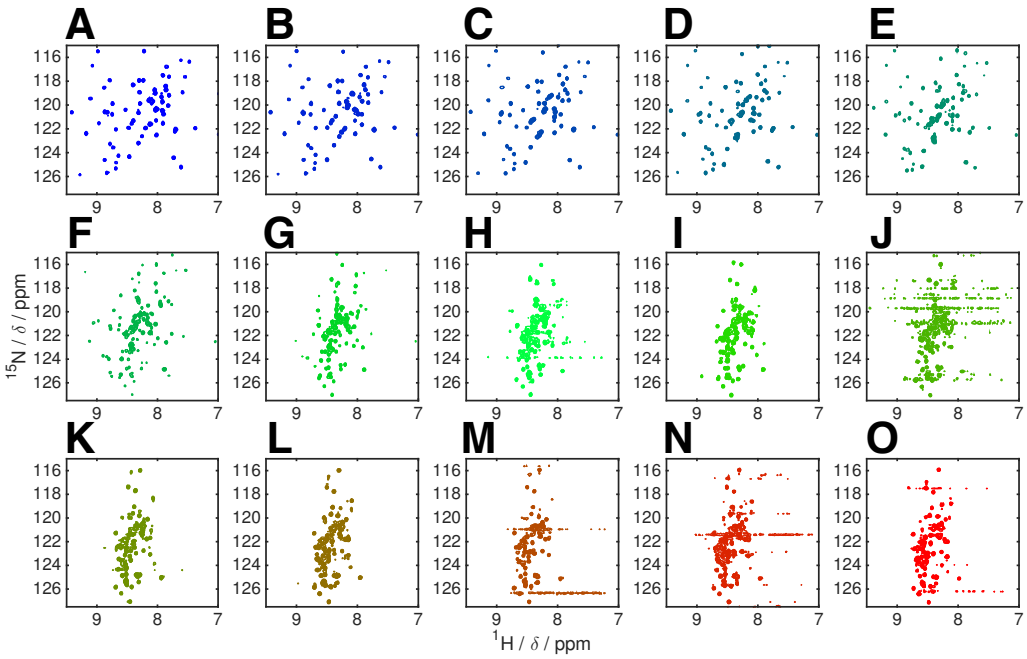

Figure 3.5: *2D spectra of the urea titration of the HP36-NTL9 fusion.* Spectra were acquired at 500 MHz and 328 K. Urea concentrations are as follows: **A:** 0, **B:** 1.0, **C:** 2.0, **D:** 3.0, **E:** 4.1, **F:** 5.0, **G:** 5.9, **H:** 6.7, **I:** 7.0, **J:** 7.8, **K:** 8.0, **L:** 8.7, **M:** 9.6, **N:** 10.2 and **O:** 11.0 M.

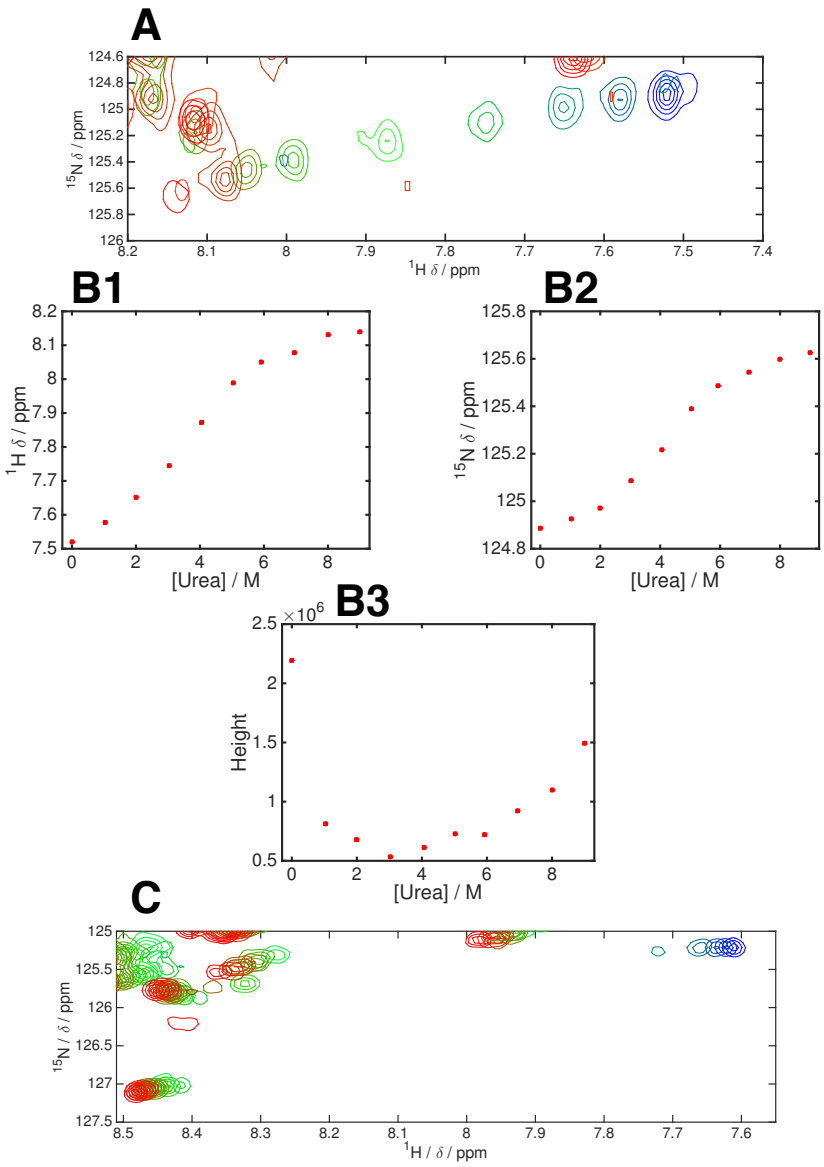

Figure 3.6: *Analysis of the 2D spectra of the HP36 urea titration.* **A:** Change in ^{1}H, ^{15}N chemical shifts and peak heights as a function of urea concentration for the A57 peak of HP36 at 328K and 500 MHz (see Figure 3.5). **B1-B3:** Changes in chemical shift and peak heights of the A57 peak with urea at 328K. **C:** Appearance of the A57 peak over the course of the titration at 298K, clearly showing the line-broadening associated with slow exchange. All spectra were acquired at 500 MHz.

of such states is often associated with substantial NMR line-broadening effects.

Another means of resolving the discrepancy in kinetics observed here is to consider the impact of urea and temperature upon kinetics. The ultrafast folding rates determined for both the GA module and the HP36 were observed using T-jump methods. In the case of the GA, it may be that the folding rate only applies at the extremely elevated temperatures used. However, in the case of the HP36, although the T-jump measurements were carried out at elevated temperatures, a backward extrapolation was made to determine the room temperature folding rate [96]. Hence, temperature alone does not appear to explain this discrepancy. It is possible then that the addition of urea also has an impact on the kinetics of the HP36, causing it to fold and unfold at a much slower rate. This is a known effect of urea, and as a result of this, it is very difficult, with just a urea NMR titration, to determine the kinetics of folding of the protein in the absence of urea.

Although the kinetics of folding of the HP36 appeared to be slower than was expected, what is clear is that the folding kinetics of the HP36 appear to be less susceptible to alteration by the aqueous environment than the GA module, since fast exchange could be observed for the headpiece (at 328 K), but not for the GA. It is also clear that the HP36 readily adopts a partially folded conformation at lower concentrations of urea, and lower temperatures than the GA. This difference in stability of the peptides may also be an interesting avenue to explore on the ribosome.

Given that the HP36 possesses not just an ultrafast folding rate, but also a folding rate that is relatively 'stable' against mutation and some of the effects of denaturants such as urea, this peptide might still be considered the most likely candidate to display fast exchange behaviour on the ribosome.

3.5 Ribosome binding experiments

Protein interactions with the ribosome can lead to line-broadening of resonances as a result of greatly enhanced relaxation rates, rendering them difficult or impossible to observe. Isolated samples of ^{15}N labelled GA module and HP36 in Tico buffer were mixed with ribosomes and HSQC spectra were recorded and compared with spectra of the free proteins, in order to assess the line-broadening effect. For the HP36, it was necessary to remove the N-terminal NTL9 tag as interactions between this protein and the ribosome would be liable to significantly interfere with results (see section 7.1.6 for the method of purification of the cleaved HP36 domain).

For both the HP36 and the GA module, it was found that inclusion of arginine in the NMR buffer was a

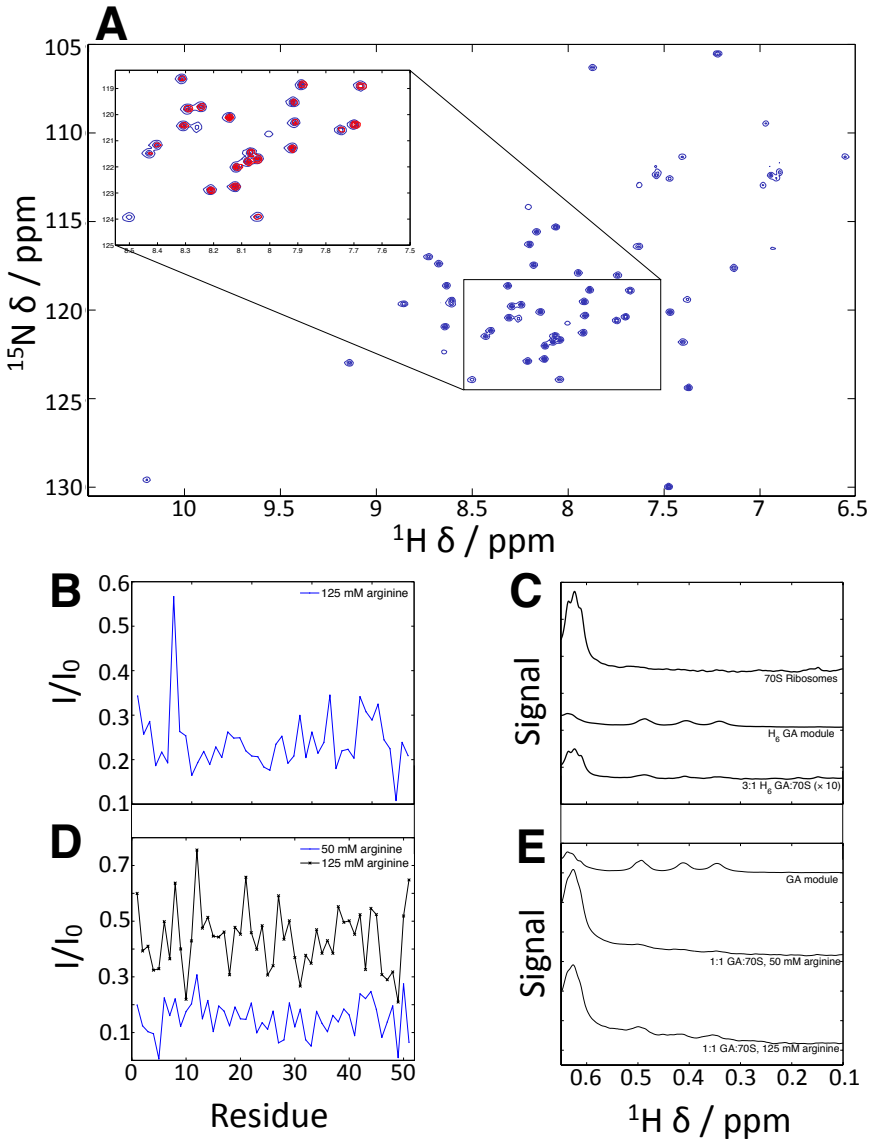

Figure 3.7: *NMR analysis of the titration of the GA module with ribosomes.* **A:** HSQC spectrum of the GA module. The inset shows the HSQC of the GA in blue overlaid with the intensity normalised HSQC of the GA module in the presence of ribosomes, in red. **B-C:** HSQC relative intensities and 1D spectra for the his-tagged GA module in the presence of ribosomes at a 3:1 protein:70S ratio (23:7.5 μM). Spectra were acquired with a 125 mM arginine supplement. **D-E:** HSQC relative intensities and 1D spectra for the non-tagged GA module in the presence of ribosomes at a 1:1 protein:70S ratio (7 μM), and an arginine supplement of either 125 or 50 mM. Plot D also shows the effect of arginine on relative intensities. All spectra were acquired at 700 MHz and 298 K.

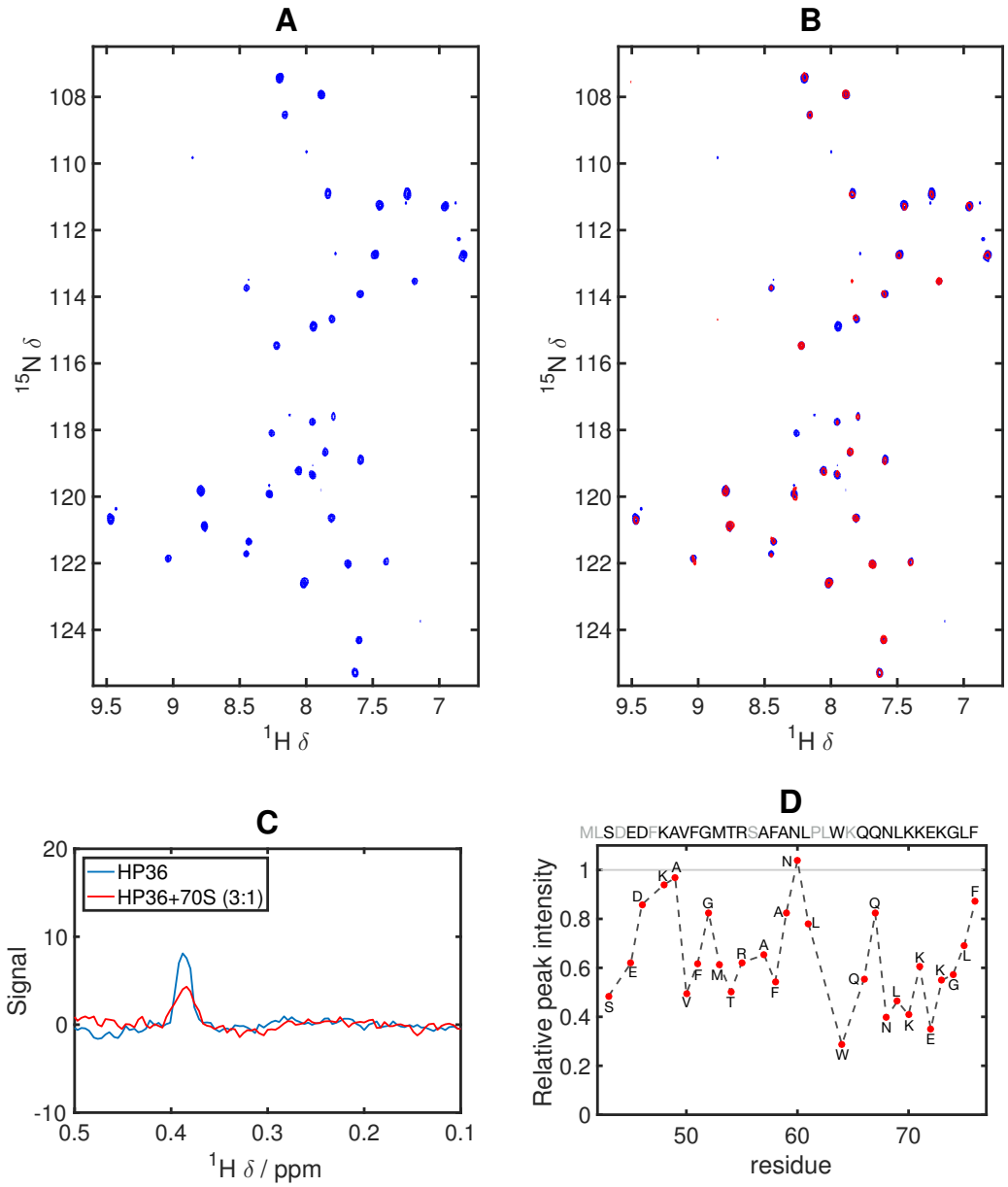

Figure 3.8: *NMR analysis of the titration of the HP36 domain with ribosomes.* **A:** HP36 SOFAST HMQC spectrum **B:** HP36 spectrum overlaid with HP36 in the presence of ribosomes (HP36:70S = 2.6:1 (6.6:2.5 μM)). **C:** 1D spectra of the HP36 with and without ribosomes (2.6:1 ratio). **D:** Relative ^{1}H-^{15}N HMQC intensities of the HP36 in the presence of ribosomes. The full HP36 sequence is shown above the plot (grey shaded residues indicate absence from plot). Each plot point is also labelled with the amino acid it corresponds to. Spectra were acquired at 700 MHz and 298 K in Tico buffer with a 100 mM arginine supplement.

necessary step for effective experiments. For the GA module, precipitation occurred as soon as the protein made contact with ribosomes, unless arginine was included, and arginine was also required to prevent HP36 line-broadening in the presence of ribosomes from eliminating signal entirely. Arginine is well known as an additional buffer component for suppressing the aggregation of proteins without affecting tertiary structure [110]. It is generally held that the weak binding of arginine ions to protein surface aid in solubilisation of the protein [111, 112], and it has also been proposed that interaction between the guanidinium group of the arginine with aromatic rings of the protein [113, 114] aids in solubilisation by screening hydrophobic patches.

Line-broadening effects were quantified by comparing the relative intensity of HSQC/HMQC peaks in the presence of ribosomes to those of the isolated proteins. The GA module interacts very strongly with the ribosome, and even forms a precipitate unless appropriate conditions are used. By including arginine in the NMR buffer, it was possible to partially recover the 2D and 1D ^{1}H signal. See Figure 3.7.

The HP36 also exhibits substantial line-broadening in the presence of ribosomes. However, with the addition of arginine, this line-broadening effect was mitigated sufficiently to allow the recording of SOFAST HMQC spectra. Normal HSQC spectra could not be observed, however (see Figure 3.8).

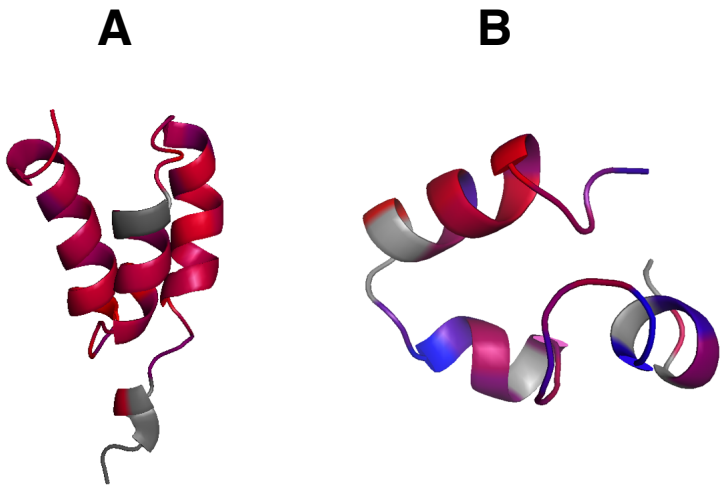

Figure 3.9: *Structures of the GA module and HP36, coloured by residue according to extent of line-broadening in presence of ribosomes.* **A:** GA module. **B:** HP36 villin headpiece. For both proteins, C-termini are at the top. Red indicates a high degree of line-broadening, while blue indicates less broadening. Grey regions indicate no data was available.

Residue specific line-broadening

Figure 3.9 shows the colour map of line-broadening of the HP36 and GA module by residue. It is difficult with this information to pick out a specific pattern of broadening in terms of locality, or secondary/tertiary

structure for the GA module. However, in the case of the HP36, the C-terminus appears to display a slightly greater degree of line-broadening than the N-terminal end.

3.6 Evaluation of the main candidates

Of the two proteins investigated here, the HP36 appears to be the stronger candidate for the production of a nascent chain sample. It is important to stress here, that the use of urea to determine folding kinetics may lead to an underestimation of folding kinetics. Other folding studies of the HP36 indicate that it tends to retain a high folding exchange rate even when destabilising mutations are introduced [97]. The effect that the ribosome has on kinetics is essentially impossible to predict at this stage, and so the only option is to produce nascent chain samples in order to investigate the behaviour of any protein in the ribosomal environment.

Also, while interactions with the ribosome appear strong in both proteins, this should not rule out their further development. Firstly, it is possible with arginine, and possibly other additive components, to mitigate this interaction in order to maximise observable signal, and secondly, because such interactions may in fact be a crucial aspect of co-translational folding in some cases. Hence, to ignore such interactions, or to only investigate proteins where no interaction is observed, may exclude an important aspect of co-translational folding from our research. As such, both the proteins investigated here were considered suitable candidates for study as nascent chains.

3.7 NMR resonance assignments

Once it was clear that the HP36 and the GA module were both suitable for consideration as fast folding nascent chains, resonance assignments were obtained for the ^{15}N HSQC spectra. The ^{15}N resonance assignment of the HP36 headpiece domain was taken from the biological magnetic resonance data bank (BMRB, entry 11519 [115]).

Since there was no assignment available for the GA module in the literature, this protein was assigned using a combination of triple resonance techniques. A uniformly ^{15}N-^{13}C labelled sample of the GA module with an N-terminal TEV cleavable H_6 tag was expressed and purified using nickel column chromatography and size exclusion. The full sequence of the expressed protein was MSHHHHHHSSGENLYFQGTIDQWLL KNAKEDAIAELKKAGITSDFYFNAINKAKTVEEVNALKNEILKAHA (TEV cleavage site is underlined).

The triple resonance spectra used to carry out the assignment of the GA module were as follows; Constant

time ^{13}C HSQC with delays of 13.3, 26.6 and 40 ms, BEST-HNCO, BEST-HNCACB, BEST-HNCOCACB, (H)CC(CO)NH, H(CC)(CO)NH and HC(C)H TOCSY. This allowed for the assignment of 49 peaks of the 51 that were observed in ^{1}H-^{15}N HSQC spectrum of the GA module. This corresponds to an assignment of 96 % of the observable peaks, or 92 % of the total number of GA module residues (53). For details of the full assignment, see Appendix B.

3.8 Chapter summary

To conclude this chapter, it was shown that even supposedly fast folding proteins were in slow or intermediate exchange on the NMR timescale. The best candidate in terms of folding rate is clearly the HP36, as it maintains a rapid folding rate even at moderate temperatures. The line-broadening shown by the HP36 in the presence of ribosomes can be mitigated by including arginine in the buffer, and the exploration of charge varying mutants of HP36 may also be helpful in reducing the extent of line-broadening observed.

Perhaps the strongest reason for proceeding with the HP36 is its constant folding rate. With the exception of the room temperature urea titration, the value of k_{ex} remains very high in the presence of denaturants, preserving the fast exchange regime. Also, partially folded mutants display similarly high folding rates. Often, when a protein is found to populate the folded and unfolded states to similar degrees ($\sim$ 50 %), the balance of k_u and k_f leads to a greatly reduced k_{ex}, and so a situation such as found for HP36 is particularly ideal. This may also explain why slow exchange was more strongly observed for the GA module, as intermediate urea concentrations may have contributed more to reducing k_f than to increasing k_u.

The line-broadening by interaction with the ribosome could be resolved by the inclusion of arginine in a nascent chain sample, and also the elimination by mutagenesis of residues that effect the interaction. However, there appears to be no consistent pattern in terms of which residue types affect the HP36-ribosome interaction (charge, hydrophobicity, etc). Instead, one could choose to target and mutate specific residues that have particularly low ^{1}H-^{15}N-HSQC relative intensities in the presence of ribosomes.

These results also indicate how challenging it is to evaluate from kinetic data alone whether or not a peptide will fold fast enough to be in the fast exchange regime. The GA module and HP36 domain should be approaching fast exchange according to observations made in the literature, however this clearly was not the case. The HP36 clearly has the particular advantage of having a folding rate that remains similar under a wide range of conditions. The advantage afforded to the HP36 by this property makes it a stronger candidate for consideration as a nascent chain, and so this peptide was initially selected for further development.

We now discuss work made towards developing these proteins into DNA constructs that can be translationally stalled, allowing the for study of their structure and folding by NMR whilst attached to the ribosome.

Chapter 4

Production of Nascent Chain DNA Constructs, and Analysis by NMR Spectroscopy

The 17 amino sequence FSTPVWISQAQGIRAGP, isolated from the bacterial SecM protein, is known to arrest peptide translation [39, 116], and as such it has been exploited for a vast range of co-translational studies [49]. After incorporating the sequence into the C-terminus of the desired protein, the protein can be expressed under similar conditions to the non-SecM variant, with typically much shorter expression times (around 30 minutes at 30 °C for *E. coli* cells in the mid-exponential phase; see section 7.3). We now describe attempts to incorporate this motif into the DNA constructs of the HP36 and GA module domains.

4.1 Testing of expression and purification conditions for the HP36 RNC construct

Three separate DNA constructs for this nascent chain were trialled. They were produced by a combination of site directed mutagenesis and digestion/ligation techniques (Section 7.4).

The three target constructs investigated had the following sequences;

H_6-TEV-HP36-SacI-dom6+110-SecM

H_6-TEV-NheI-NTL9-Xa-HP36-SacI-dom6+110-SecM

H_6-NTL9-TEV-HP36-SacI-dom6+110-SecM

The constructs were termed HP-RNC1, HP-RNC2 and HP-RNC3, respectively.

4.1.1 Expression testing of RNC constructs

Transformed plasmids were grown in LB media (50 ml in 200 ml conical flasks) at 37 °C and 250 rpm to an optical density of 0.6. Expression was induced by addition of 1 mM IPTG and cultures were then incubated either at 37 °C for 1 hour, 30 °C for two hours, or 20 °C for 16 hours. In all cases, reproducible expression of a nascent chain as detected by immunoblotting could not be obtained.

The expression attempt was repeated by growing cells to saturation in MDG medium before transferring to EM9 and inducing expression under the same conditions. Again, reproducible expression of sufficient yield could not be obtained. Conditions were found under which the HP-RNC3 construct did express, however this could not be reproduced, despite numerous attempts. It was therefore concluded that further work is required to make the system more robust before proceeding with development of this nascent chain.

4.2 Designing and cloning the GA module RNC DNA construct

Since the HP36 proved extremely difficult to express, we subsequently focussed on developing conditions for expression of the GA module nascent chain, a generally more robust protein. These attempts proved successful (Section 4.3.1). Using the FLN750 RNC work as a starting point, it was decided that the first study would be ^{15}N labelled, and that translation should be stalled when the GA is as far away from the ribosome as possible in order to ensure that the structured form of the peptide be observed, since this is likely to be easier to identify. The longest Ig2 linker that has been studied places 110 amino acids between the PTC and the protein of interest, and so this linker was incorporated into the RNC construct.

To produce the required construct, NdeI and SacI digestion sites were introduced by PCR into the Ig2+110 RNC and GA module DNA constructs. The plasmids were then digested with NdeI and SacI enzymes, and the GA insert and Ig2 vector were isolated by gel electrophoresis and excision, and recombined with T4 DNA ligase (Figure 4.1A). The identity of the plasmid after ligation was confirmed by DNA sequencing (Source BioScience). The expected mass of the amino acid construct (including the his-tag) is 19.9 kDa. For details of DNA manipulation, see section 7.4. The DNA plasmid hence used for expression of the GA module nascent chain had the following sequence;

H_6-TEV-GA-dom6+110-SecM.

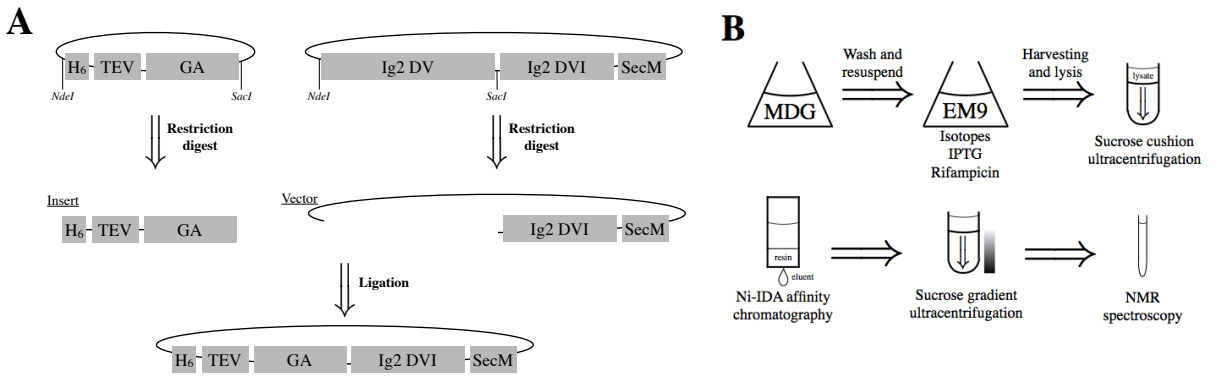

Figure 4.1: *Schematics outlining the process of producing a translationally stalled protein.* **A**: Production of the RNC DNA construct used in the present report. **B**: Schematic of protocol for preparation of an isotopically labelled RNC.

4.3 Expression and purification conditions for the GA module RNC construct

The GA module RNC was found to express well under the same conditions used for the FLN750 RNC (vide infra). Attempts were made to cleave the his-tag from the GA module using TEV protease, however the nascent chain could not be recovered after incubation with the protease, suggesting either excess proteolysis or aggregation had occurred. Hence, it was determined that initial NMR studies on this nascent chain would be carried out with the his-tag intact.

4.3.1 Expression and purification of the ^{15}N labelled nascent chain

Figure 4.1B summarises the protocol for expression and purification of isotopically labelled nascent chain complexes. The culture was initially grown in an unlabelled non-inducing MDG medium to maximise the cell growth. Once cell growth reached saturation ($OD_{600} \approx$ 4-6) the cells were transferred to a minimal EM9 medium (M9, pH 8.00) in the presence of the desired isotopic components ($^{15}NH_4Cl$). As soon as the cells were resuspended, expression was induced with 1 mM IPTG for 10 minutes at 37 °C and 200 rpm. After 10 minutes, rifampicin was added (150 mg/L), and incubation continued for a further 35 minutes. The cells were immediately chilled on ice, before harvesting at 4000 rpm for 20 minutes.

The expression time is optimal for maximising the extent to which the ribosomes in the culture are occupied with stalled nascent chains and minimising the extent of background isotopic labelling and nascent chain

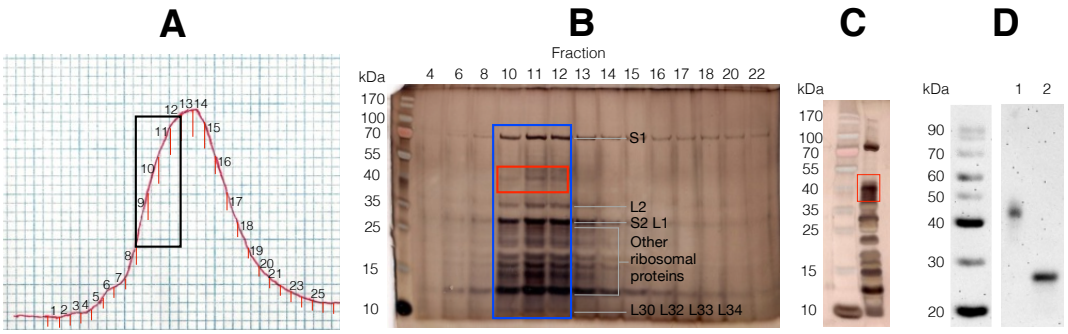

Figure 4.2: *Summary of the purification of the GA module RNC.* **A:** Elution of ribosome bound nascent chains from a 10-35 % sucrose gradient. The highlighted fractions correspond to 70S ribosomes. **B:** Silverstain gel of the indicated fractions from the gradient. Fractions containing 70S ribosomes are boxed in blue, and the red box highlights background due to impurities. The pattern of bands observed in the red box is consistent with denatured ribosomal proteins. Major bands are labelled according to their corresponding ribosomal protein. **C:** Silverstain of final purified sample (9 pmol) of the fractions pooled from B. **D:** Western blot of final sample. The lane adjacent to the marker was loaded with a low pH dye, and the far lane was loaded at high pH after incubation with RNase for 5 minutes to ensure release of the nascent chain. Both lanes contain 9 pmol of ribosomes.

release. The rifampicin also minimises background labelling by inhibiting the activity of RNA polymerase [117], hence preventing further ribosome growth in the labelled medium, which would lead to non-specific ribosome labelling.

After expression, cells were lysed by french press, and the soluble fraction loaded onto a sucrose cushion to extract the ribosomal component. The pellet contained ribosome particles of various sizes. The nascent chain bound ribosomes were purified using an IDA-Nickel resin, eluting with 150 mM imidazole, followed by further purification of the 70S particle by sucrose gradient ultracentrifugation (Figure 4.2A). The pure 70S containing fractions were identified by silverstained SDS-PAGE (Figure 4.2B), pooled, concentrated and exchanged into Tico buffer (5×, 15 ml to 0.5 ml) before being snap frozen in liquid nitrogen and stored at -80 °C. The TEV cleavage site was left intact for this sample, and arginine was also excluded from the Tico buffer.

4.4 Biochemical assessment of the nascent chain

The presence of the stalled nascent chain was confirmed by anti-his western blot analysis. The stalled translation product is covalently bound to tRNA at the PTC, hence the attached nascent chain runs approximately 20 kDa higher than expected for the protein alone (Figure 4.2D, lane 1). Addition of RNase to a sample prior to loading the gel (Figure 4.2D, lane 2) degrades the tRNA, releasing the protein such that it migrates

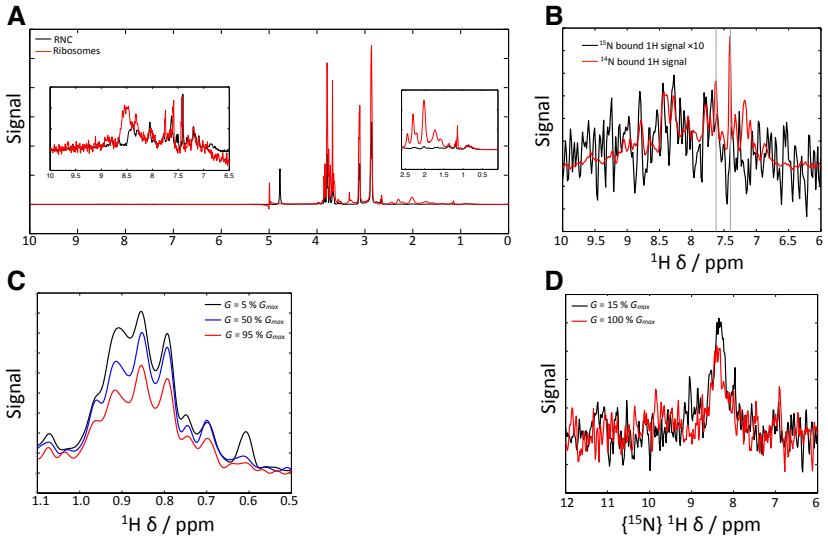

Figure 4.3: *Evaluation of initial sample integrity by NMR.* **A:** 1D ^{1}H spectrum of the GA RNC in black, overlain with a ribosome spectrum in red. Signals were normalised for concentration and number of scans. Insets show zoomed in regions of interest, namely the methyl and amide proton regions. **B:** ^{15}N edited (black) and ^{14}N edited (red) 1D ^{1}H spectra of the nascent chain sample. The ^{15}N edited (black) spectrum indicates the ^{15}N bound ^{1}H signal arising from the nascent chain sample, while the ^{14}N edited (red) spectrum indicates the ^{14}N bound ^{1}H signal arising from the same sample. The intensity of the ^{15}N edited spectrum was multiplied by 10 for comparison with the ^{14}N edited signal. The grey vertical lines on the spectra indicate the presence of peaks corresponding to ribosomal proteins L7/L12. These were used to ensure that the spectra were normalised relative to ribosomes, rather than nascent chain. Since the ^{15}N bound signal is ten times weaker than the ^{14}N bound signal, this indicates that the extent of background ^{15}N labelling of the ribosome is approximately 10 %, which is sufficiently low to allow ^{1}H-^{15}N analysis of the nascent chain itself. **C:** Monitoring of the methyl ^{1}H signal in the STE diffusion experiments at gradient strengths of 5 % (black), 50 % (blue) and 95 % (red) of G_{max}. **D:** Monitoring of the amide ^{1}H signal in the SORDID diffusion experiments at gradient strengths of 15 % (black) and 100 % (red) of G_{max}.

down the gel at its actual mass. Although it was found that the tRNA released form migrates around 7 kDa higher than expected (27 kDa instead of 20), this was attributed to the highly disordered nature of the domain 6 linker, which can lead to a higher hydrodynamic radius during gel migration. Hence, this confirmed that translation had been successfully stalled. The purity of the 70S particles was evaluated by silverstained SDS-PAGE (Figure 4.2C), indicating that the sample was of high purity.

To determine occupancy, the isolated GA module was purified and used as a control to measure the concentration of bound nascent chain in the RNC sample by western blotting with densitometry analysis. However, the isolated GA module did not appear to bind to the anti-his antibody, and so could not be used as a control. It is uncertain as to why the isolated protein does not bind the antibody whilst the RNC construct does.

4.5 NMR analysis of the GA module nascent chain

The frozen sample was thawed on ice, and exchanged into fresh Tico buffer 3 times to filter out nascent chain that can release in the snap freeze/thawing process. The final NMR sample (including 10 % D_2O and 0.01 % DSS) contained 9 μM ribosomes with a volume of 320 μl.

4.5.1 NMR spectra

All spectra were recorded at 25 °C on a 700 MHz Bruker Avance spectrometer equipped with a cryogenic probe. ^{15}N filtered and ^{15}N edited spectra were initially recorded to estimate the extent of background ribosome labelling, and the nascent chain signal was estimated from a ^{15}N edited 1D ^{1}H spectrum. The sample was then analysed by recording ^{15}N SOFAST HMQC correlation spectra, ^{15}N edited SORDID diffusion, ^{1}H-stimulated echo (^{1}H-STE) diffusion, and ^{1}H 1D spectra in an interleaved manner for the duration of the sample lifetime (27 hours in total), allowing its behaviour to be monitored systematically over time.

SOFAST HMQC spectra were recorded with 1024 time points in the direct dimension and 64 points in the indirect (^{15}N) dimension. The centre of the spectrum in the indirect dimension was set to 117 ppm with a sweep of 32 ppm, corresponding to a resolution of 0.5 ppm (35.5 Hz). This represents a minimisation of resolution in order to maximise the signal obtained, considering this is an untested nascent chain and the initial signal was weak relative to that from previous spectra recorded on the domain 5 nascent chain (vide infra).

4.5.2 Evaluation of initial quality and integrity

1D spectra

Inspection of the 1D ^{1}H spectrum reveals a profile characteristic of purified ribosomes, as shown by overlaying the RNC 1D spectrum with a 1D spectrum of previously purified ribosomes (Figure 4.3A).

Evaluation of background isotopic labelling

Effective NMR studies on nascent chains require a high percentage of isotopic labelling of the nascent chain with minimal to no labelling of the ribosome. Even small amounts of signal arising from ^{15}N labelled ribosomes can interfere with observed data. To evaluate the extent of background labelling, 1-dimensional ^{15}N edited and ^{15}N filtered (or ^{14}N edited) spectra were recorded on the nascent chain sample, and the intensities of the amide region of the L7/L12 ribosomal stalk region were compared between spectra (see

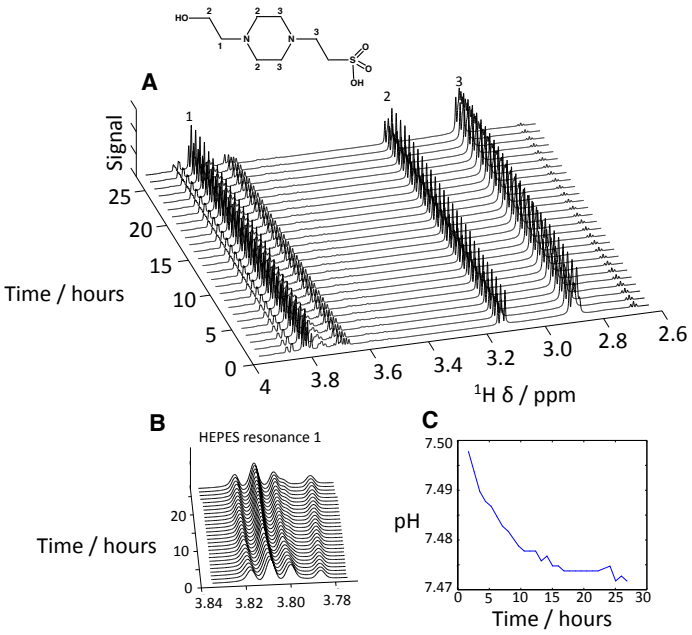

Figure 4.4: *Sample buffering capacity by monitoring of HEPES chemical shifts over time.* **A:** HEPES region of 1D ^{1}H spectrum, alongside crude assignment. **B:** Focus on HEPES resonance 1 shows subtle changes in peak position over time. **C:** Sample pH calibration over time, determined by comparing the middle peak from the HEPES resonance 1 triplet with previously recorded standards.

Figure 4.3B). The two peaks at 7.4 and 7.6 ppm in the ^{15}N filtered spectrum correspond to ^{14}N bound protons in the L7-L12 stalk. By scaling this to the intensity of the same region in the ^{15}N edited spectrum, the extent of ribosome ^{15}N labelling can be seen to be approximately 10 %. The ^{15}N edited and ^{14}N edited spectra are collected by initially exciting all amide protons in the sample, followed by a magnetisation transfer step onto ^{15}N or ^{14}N depending upon which nucleus we intend to edit for. Magnetisation is then transferred back to ^{1}H for detection. In each case, the pulse sequence is kept as similar as possible, so that intensity differences can be attributed to differences in the relative quantities of either nucleus.

Diffusion

Although the presence of the tRNA bound form of the nascent chain had already been confirmed (*vide supra*), NMR diffusion is routinely used to confirm that 1: the ribosomes are intact and 2: the tRNA bound nascent chain is still part of the 70S complex. To accomplish this, a ^{1}H stimulated echo (^{1}H-STE) diffusion experiment was used to determine the diffusion coefficient of ribosome methyl protons in the 0.5-1 ppm region of the 1D spectrum, and a ^{15}N XSTE SORDID experiment [51] can monitor in the amide proton region (7-10 ppm), to selectively measure nascent chain diffusion, assuming minimal background labelling.

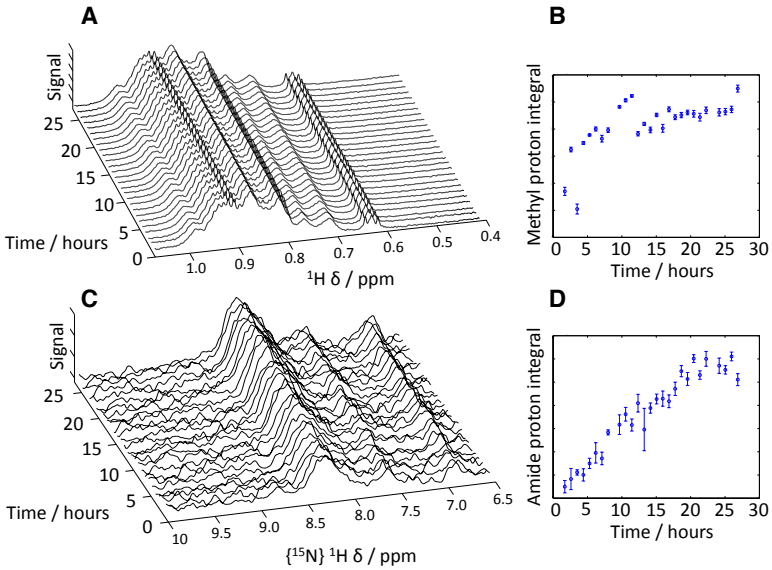

Figure 4.5: *Evaluation of changes in signal over time to monitor sample breakdown.* **A:** Methyl proton region of the 1D ^{1}H spectrum. **B:** Integrated signal from the methyl ^{1}H region of the 1D spectra. **C:** Amide ^{1}H region from the first increment of the SOFAST HMQC spectra. **D:** Integrated signal from the amide ^{1}H region of the SOFAST spectra.

For quantitative evaluation, the integrated signal of the selected regions of the spectra was used to calculate diffusion coefficients using the Stejskal-Tanner equation [118],

$$\frac{S}{S_0} = \exp\left\{-\gamma^2 G^2 \delta^2 \sigma^2 \left(\Delta - \frac{\delta}{3}\right) D\right\} = \exp\{-BD\} \tag{4.5.1}$$

where γ is the ^{1}H gyromagnetic ratio, G is the strength of the encoding and decoding gradients, δ is the gradient duration, Δ is the diffusion delay between gradients, σ is a shaped pulse factor ($= 0.9$), and D is the translational diffusion coefficient, in m^2 s^{-1}.

For the ^{1}H-STE pulse sequence, three gradient strengths, at 5, 50 and 95 % of G_{max}, (where $G_{max} = 0.565$ T m^{-1}) were used. In the interests of time, only 3 gradient strengths were recorded, which is too few to obtain a high quality fitted curve. Instead, D was obtained from the ratio of the signal at gradient strengths of 5 and 95 %, or 50 and 95 %.

For the SORDID experiments, 2 gradient strengths, at 15 and 100 % of G_{max} (where $G_{max} = 0.393$ T m^{-1}) were applied. Since the ^{15}N edited signal was weaker than the pure ^{1}H signal, and ^{15}N diffusion experiments take longer to run, only 2 gradient strengths were used.

The methyl and amide regions of of the 1D spectra from the first diffusion experiments are displayed in

Figure 4.3C-D. Initial diffusion coefficients indicate an intact, attached, ribosome bound nascent chain (See Figure 4.6).

4.5.3 Monitoring sample integrity over time

To determine the window in which an attached nascent chain was clearly being observed without contamination from released translation products, the diffusion and 1D spectra were used to watch the sample as it broke down over time.

The stability of the buffer was monitored by observation of HEPES chemical shifts in the proton spectra (Figure 4.4). Particularly, the HEPES resonance arising at around 3.8 ppm is used to monitor sample pH, and a shift in this resonance has previously been correlated with breakdown of nascent chain samples. In Figure 4.4B-C, the majority of HEPES resonances appear relatively unperturbed over time, corresponding to a very minor change in pH of 0.02 over the 27 hour period that data was recorded.

The ribosomes were monitored by observing STE diffusion in the methyl proton region over time, as well as by looking at the overall integral of this region from the proton spectra. As ribosomes gradually fall apart, the integral of the methyl region increases to reflect reduced correlation time (Figure 4.5A-B), while the diffusion coefficient D increases. Interestingly, the value of D is greater when calculated using G = 5 and 95 % as opposed to 50 and 95 % (Figure 4.6A), which is due to the former measurement being biased towards smaller molecules that are detected in the 5 % signal, but not at 50 or 95 %. A diffusion coefficient between 1.7 and 2.2 m^2 s^{-1} is usually considered an appropriate range for intact ribosomes, and this was observed for at least 12 hours, and perhaps up to 22 hours, suggesting good quality ribosomes for the majority of the experimental time.

Although the ribosomes were intact for at least 12 hours, it was also important to monitor ^{15}N edited signal to confirm that the nascent chain was attached. Figure 4.5C-D shows the ^{15}N edited amide signal over time, which was determined by only processing the first increment of SOFAST HMQC spectra. ^{15}N diffusion coefficients were calculated from the SORDID results, (Figure 4.6). Since the ^{15}N signal was very weak, sequential experiments had to be summed in order to obtain sufficient signal that reliable diffusion coefficients could be obtained. The SORDID diffusion highlights that although the ribosomes were intact for at least 12 hours, the ^{15}N diffusion coefficient rose above 2.2 m^2 s^{-1} after 2.5 hours of recording time, meaning that only the first 2.5 hours could be considered to be reporting exclusively on the intact, translationally stalled protein.

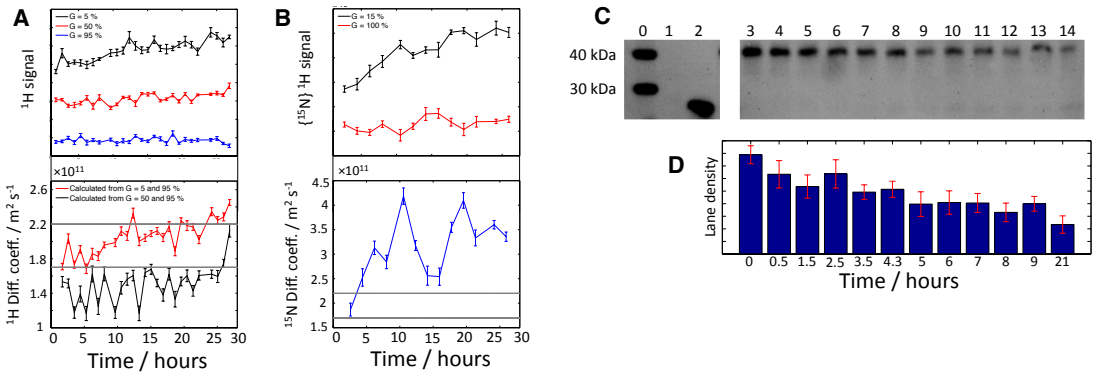

Figure 4.6: *Further sample monitoring by measurement of translational diffusion coefficients over time.* **A:** Above: Methyl proton signal from STE diffusion experiments at gradient strengths of 5, 50 and 95 %. Below: Diffusion coefficients calculated from STE results, using either $G = 5$ and 95 % (red), or $G = 50$ and 95 % (black). **B:** Above: Amide proton signal from SORDID diffusion experiments at gradient strengths of 15 and 100 %, after summation of adjacently recorded spectra to improve signal to noise. Below: Diffusion coefficients calculated from SORDID results. Horizontal lines in the diffusion coefficient plots indicate the expected range for intact 70S ribosomes. **C:** Anti-his western blot analysis of RNC sample in parallel with SORDID diffusion experiments. Lanes 0, 1 and 2 correspond to marker, 70S ribosomes and the RNase treated form of the nascent chain sample, respectively. Lanes 3 through 14 correspond to aliquots of the intact RNC frozen at incremental time points (5 pmol of ribosomes per lane). **D:** Densitometry analysis of the western blot, showing a gradual reduction in lane density over time as the attached nascent chain is released/degraded.

4.5.4 Parallel biochemical experiments

Alongside the SORDID diffusion, the sample was also monitored by a parallel western blot analysis. As the sample was prepared for NMR, a small aliquot of 400 pmol was taken and kept under exactly the same conditions as the main sample. This was divided into further 5 pmol aliquots which were frozen at the same time each SORDID experiment was collected, allowing for parallel anti-his western blots to be carried out at a later date (Figure 4.6C). Densitometry analysis of the western blot was performed using the imageJ software, which revealed the intensity of the tRNA bound form to be reducing over time due to release and proteolysis. These results also indicate, along with the SORDID measurements, that the sample remained intact for at most 2.5 hours. After this, the gradual reduction in density suggests that some form of release was beginning to occur, thus contaminating bound nascent chain signal with released nascent chain signal.

4.5.5 2 dimensional correlation spectra of the RNC

The combined analysis of ^{15}N edited 1D and diffusion experiments, and parallel western blots, seems to indicate that the nascent chain remains attached for up to 2.5 hours. After this, we can expect subsequent

data to be reporting on released translation products, which therefore interfere with remaining attached protein.

Three ^{15}N SOFAST HMQC spectra were collected in the initial 2.5 hour time period, and their summation is shown in figure 4.7A. Comparing this spectrum with that of the analogous domain 5 RNC (+110 linker, Figure 4.7E), shows that the GA RNC has significantly weaker signal, which may be due to a low 70S occupancy of the nascent chain, as well interactions between the GA and the ribosome surface.

To increase the signal, all of the SOFAST spectra recorded in the 27 hour period (Figure 4.7B) were added together, allowing for more facile analysis of the origin of the ^{15}N signal. Numerous dispersed peaks arise in this summation, most of which overlay very well with the L7/L12 stalk (4.7C), and the general outline of the RNC signal appears to fit well to either L7/L12 or the domain 5 RNC (4.7C and 4.7G). However, a group of well resolved resonances at around 8.2, 112.7 ppm was observed, which are a known marker for the domain 6 linker (Figure 4.7G). This therefore confirms the presence of the nascent chain. Although this signal could not be observed in the time period that the peptide was attached, this helps confirm the identity of the sample. No conclusive overlay could be observed with the isolated GA module at any stage (4.7D), indicating perhaps that low nascent chain concentration, along with protein:ribosome interactions, rendered GA signal too weak or broad to be detected.

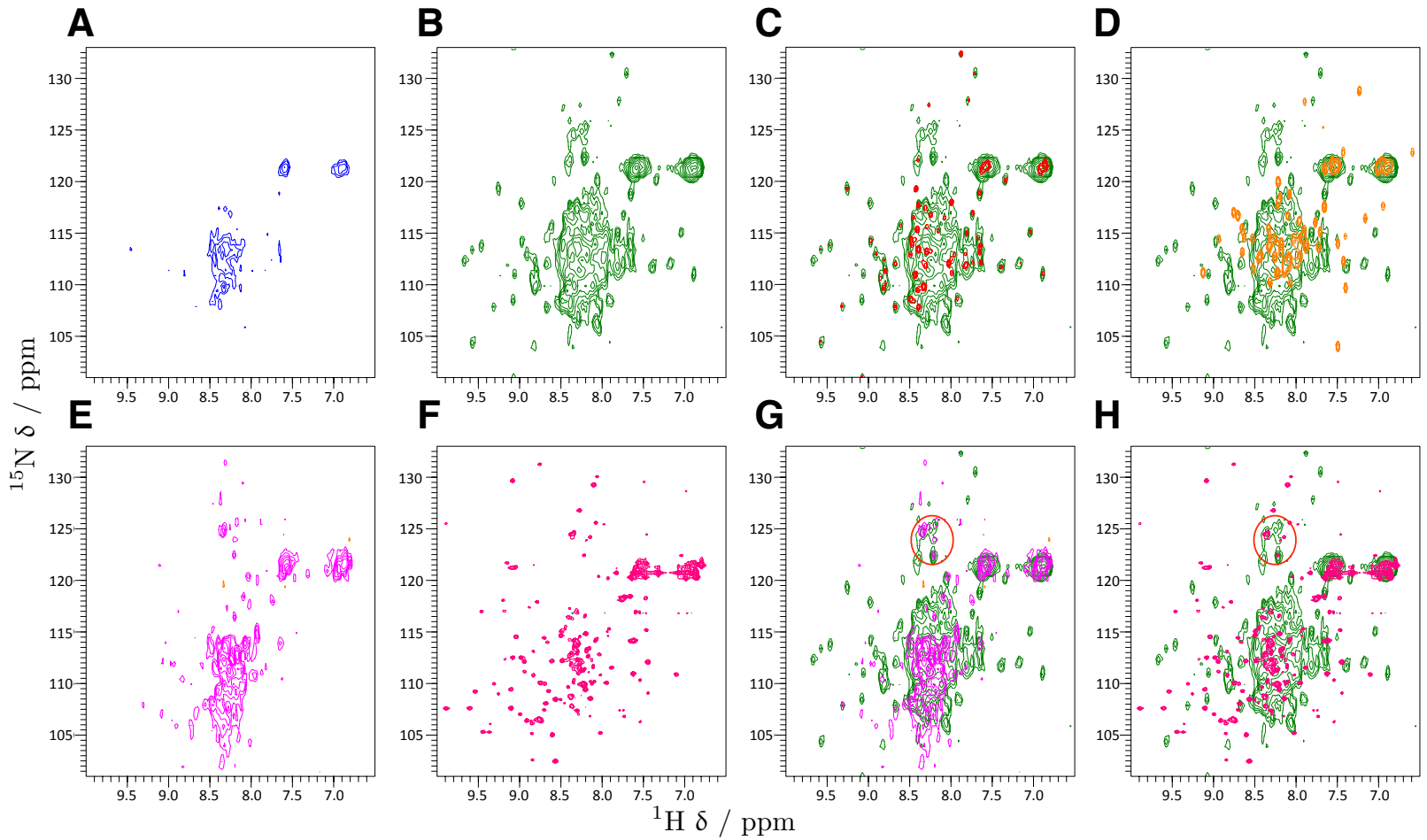

Figure 4.7: *Analysis of 2D spectra of the GA module nascent chain.* **A:** ^{1}H-^{15}N SOFAST-HMQC spectrum of the GA module RNC. 768 scans, summed over 3 spectra from the first 2.5 hours of NMR time. **B:** Summation of all SOFAST spectra collected on the GA RNC over 27 hours. **C:** Overlay of B with an HSQC of the isolated L7/L12 bacterial ribosomal protein, showing good overlay with the dispersed peaks from the RNC sample. L7/L12 spectrum was provided by Dr Xiaolin Wang. **D:** Overlay of B with a SOFAST-HMQC of the isolated GA module, showing no significant overly with the dispersed peaks. **E:** HSQC spectrum of an isolated sample of the Ig2 (domain 5 + truncated domain 6) protein. **F:** SOFAST-HMQC spectrum of the Ig2 +110 RNC. E and F were provided courtesy of Anaïs Cassaignau. **G:** Overlay of B and E. The region circled in red corresponds to an overlay between the dom5 RNC and the GA RNC. These overlaying peaks correspond to the domain 6 linker, confirming the presence of the nascent chain, albeit most likely coming from released nascent chain. **H:** Overlay of B and F, again showing some minimal degree of overlay near the centre of the spectra (circled in red). The NMR spectrum of the L7/L12 stalk was provided by Dr Xiaolin Wang. Plots were produced using the ccpn NMR analysis tool [119].

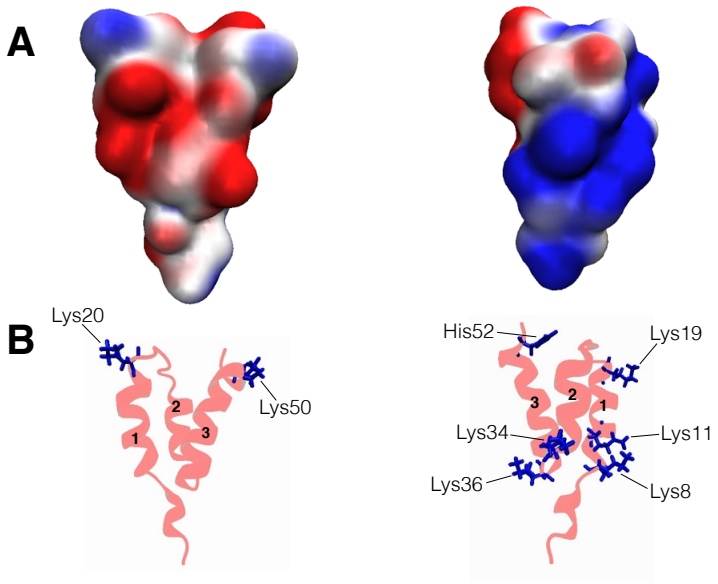

Figure 4.8: *Charge distribution of the surface of the GA module.* **A:** Charged surface of the GA module from two separate angles. Red $\Longrightarrow$ negative, blue $\Longrightarrow$ positive. **B:** tertiary structure of the GA module, highlighting the location of the positive residues. The positively charged region is largely composed of solvent exposed lysine residues. The surface charge of the GA module was estimated using the PDB2PQR web server [120].

4.6 Discussion

Although the GA module was not observed at the end of the ribosome tether, it was possible to observe signals corresponding to the disordered domain 6 linker in spite of the weak signal. Diffusion measurements may potentially have suffered from interference from background labelling of the ribosome, however the attachment of the nascent chain was also confirmed by western blot analysis, and so the current strategy now relies on manipulating the properties of the system and its experimental conditions such that discrete resonances of the attached GA module nascent chain become observable.

There are numerous factors that deserve attention in this regard. Perhaps the most pressing issue is that of nascent chain occupancy. As of yet, the occupancy of the RNC sample is still not known, although this is an important factor in determining the true concentration of nascent chain being observed. Occupancy determination will require a construct with similar affinity for the anti-his antibody as well as similar membrane transfer efficiency for western blotting. Purification of the released form of the nascent chain, or a similar sized construct, will be necessary for this purpose.

The presence of arginine in the Tico buffer appears to screen the interaction between the GA and the

70S particle, and so its presence in the nascent chain sample may be crucial for preventing extensive line-broadening. However, high arginine concentrations may require extensive pH balancing with HCl, potentially leading to higher ionic strengths than are suitable for NMR. Hence, experiments to determine the maximum arginine concentration at pH 7.5 that still provides a tenable environment for NMR should be carried out. Further investigations on the impact of the histidine purification tag also need to be carried out.

There is a possibility that electrostatics play a role in protein-ribosome interactiond. The ribosome is known to have a highly charged surface, mostly negative, and the GA module has both positive and negative regions that are solvent accessible. Previous work on the impact of electrostatics of the ribosome by Knight et al. has shown that negatively charged disordered nascent chains are repelled from the 70S surface while positive chains are somewhat attracted [3]. This is also in agreement with the recent work by Deckert et al. that has shown a similar result with α-synuclein, both as a nascent chain and as a protein:ribosome interaction study. In this case, by replacing positively charged lysine residues on α-synuclein with the negative glutamic acid, the interaction between the α-syn nascent chain and the ribosome surface was weakened, as evidenced by a lesser degree of line-broadening [53].

Figure 4.8 shows the electrostatic surface of the GA module calculated using the PDB2PQR web server [120]. The surface reveals that the exposed positive and negative residues on the GA module are largely clustered on opposite sides of the helix bundle, with most positively charged lysines residing on helices 1 and 2, whilst the negative region is clustered around helices 1 and 3. One could hypothesise that the positive patch of the GA module undergoes some form of non-specific interaction with the ribosome surface. To investigate this possibility, the gradual replacement of positively charged GA residues can be used to determine their impact on ribosomal interactions. Most of the positive residues on the GA module are lysines (Figure 4.8), and it is possible that they strongly mediate the ribosomal interaction.

To investigate the impact of positive residues on the interaction, the effect of the mutations described in Table 4.1 could be explored in terms of their impact on NMR line-broadening in the presence of ribosomes. The PoPMuSiC prediction software tool [121] was used to estimate the impact of the mutations on the overall thermodynamic stability. It was then found that since most lysines, except K11, are largely solvent exposed, the overall effect on stability of mutation to glutamic acid is relatively low in most cases, with an estimated $\Delta\Delta G$ of less than 0.2 kcal mol^{-1} for all mutations except K11E.

Therefore, one could sequentially replace solvent exposed lysines with glutamic acid, followed by determining the impact these have on the interaction between the mutant and the 70S using protein:ribosome NMR titrations. Ideally, a mutant will be found that interacts minimally with the ribosome whilst retaining

Construct	Solvent exposure / %	$\Delta\Delta G$ / kcal mol^{-1}
Wild type	N/A	0.0
K8E	63.68	-0.03
K11E	26.86	0.79
K19E	65.57	0.06
K20E	83.83	0.14
K34E	75.88	0.11
K36E	65.96	0.13
K50E	81.79	-0.08

Table 4.1: *Proposed charge replacement mutations for the GA module.* Solvent exposure and $\Delta\Delta G$ of folding were calculated using the PoPMuSiC server [121]. A $\Delta\Delta G < 0$ indicates a stabilising mutation, hence the mutations shown here are predicted to have little effect on the overall structure of the protein.

tertiary structure and a fast folding rate.

A particular concern with regards to the impact of mutation on structure and folding is that even though many of the lysines are solvent exposed and not involved in formation of the protein core, they still appear clustered opposite the negative side of the protein, and perturbation of this pattern of oppositely charged sides may adversely affect the tertiary structure. It is not easy to predict what effect will be induced however, despite indications from prediction tools, and so a systematic residue by residue test of the impact of each substitution will be necessary.

Alternately, there is a possibility that interaction between hydrophobic regions on the protein of interest and the ribosome also play a role in protein:ribosome interactions. The ability of arginine to weaken the interaction, and also to prevent precipitation of some proteins, indicates a hydrophobic contribution. The above experimental outline detailing the investigation of electrostatic contributions to the interaction could also be applied to hydrophobic contributions in much the same way, by replacing surface exposed aliphatic and aromatic residues with with non-charged polar residues.

The structural and folding impact of any mutations can be explored experimentally using circular dichroism, while NMR spectroscopy can complement this alongside being used to investigate ribosome interactions. It is hoped that the combination of appropriate mutants and high arginine concentration will screen out any interactions, whilst keeping intact the structure and rapid folding properties of the GA module. Provided this can be achieved, then steps can be taken towards manipulating the folded/unfolded state equilibrium present on the ribosome. This can be achieved through point mutations that destabilise the native state, or by reducing the linker length such that closer proximity of the GA to the ribosome may prevent complete folding of the protein.

Finally, it should be noted that the GA module and the HP36 are small peptides, which gives them

another feature that has not yet been explored by NMR. Other studies, such as in the recent work by Marino et al. [42] have shown that small peptides are capable of folding far closer to the exit tunnel than has been observed for larger peptides. The HP36 in particular was shown to display this property of folding closer to the ribosome. Hence, a finer level of structural detail on the folding of small peptides closer to the exit tunnel would be a rich avenue of research that both the HP36 and GA module would be well placed to investigate.

Chapter 5

Understanding the Role of the Ribosome Surface, and Exploiting the Fluorine Nucleus for NMR

In Chapter 3, other proteins were trialled as fast folding candidates but rejected as they were not considered suitable. Two of these include the Trp2-cage and WW domain. Both of these proteins interact strongly with the ribosome, and the WW domain in particular was found to immediately precipitate upon mixing with ribosomes. Other proteins that precipitate upon contact with ribosomes include the NTL9 and the GA module. These observations and others indicate that the ribosome has a strong effect on the environment of neighbouring proteins, and as such it must be understood in order that we can understand its biological role in protein folding.

Attempts to determine consistent factors that appear to contribute to ribosome interactions have met with limited success. For example, the his-tag on α-synuclein has a clear effect on interaction with the ribosome, as does the mutation of lysines to glutamic acid [53]. On the other hand, in the case of FLN750, no interaction is observed regardless of the presence or absence of the his-tag.

Some of the interactions between protein and ribosome may be incidental. However, the possibility that this unique environment of the ribosome has an effect on co-translational folding clearly needs to be explored. Whilst it leads to line-broadening, which renders peaks difficult to observe, it cannot be ignored as it may be a crucial component in the folding of many proteins.

In order to further our understanding of this phenomenon, we therefore investigated in greater depth the interaction between α-synuclein and 70S ribosomes. This interaction is of particular interest, since it is observed in α-syn nascent chains, and has been the subject of much detailed investigation [53].

The simplest approach towards characterisation of the interaction is a simple concentration titration. This can be modelled using well known equations for interactions between binding partners in a protein:protein interaction. The presence of the histidine tag has also been shown to have a strong effect on the interaction, leading to a significant reduction in peak intensity at the N-terminus in the presence of ribosomes. We also attempt to characterise the effect of this tag by carrying out titrations of his tagged synuclein (H_6-αsyn) in complement to the wild type.

This work also complements another avenue of research, namely the use of the fluorine nucleus for NMR spectroscopy. We show in section 5.4 of this chapter how the fluorine nucleus has the potential to provide unique insight into the nature of protein/ribosome interactions.

5.1 Characterisation of α-synuclein line-broadening in the presence of ribosomes

Using ^{15}N labelled α-syn samples and unlabelled ribosomes, we carried out ^{1}H-^{15}N HSQC T_2 relaxation experiments on α-syn at various concentrations in the presence of ribosomes.

By measurement of amide proton relaxation rates using a spin-echo HSQC pulse sequence, we are able to follow the interaction quantitatively, and show that this broadening is consistent with a simple 1:1 interaction model. This reveals a weak interaction ($K_d > 1$ mM) with extremely fast kinetics, close to the diffusion limited association rate. The residue specificity of the technique also allows identification of the area of binding, which appears localised to both the C and N-termini of the protein.

5.1.1 Sample preparation

Wild-type and polyhistidine tagged human α-synuclein constructs were overexpressed in BL21(DE3) *E. coli* cells in minimal (M9) media with $^{15}NH_4Cl$ as the sole nitrogen source. After cell lysis and extraction, H_6-αsyn purification was achieved using an NTA nickel column followed by size exclusion, while the wild type was purified by ion exchange chromatography and size exclusion on an S75 16/600 GE healthcare gel filtration column. To obtain unlabelled 70S ribosomes, BL21(DE3) trigger factor knockout (ΔTF) cells were grown in MDG medium to an OD_{600} of 4-6. We used ΔTF cells since the presence of trigger factor even

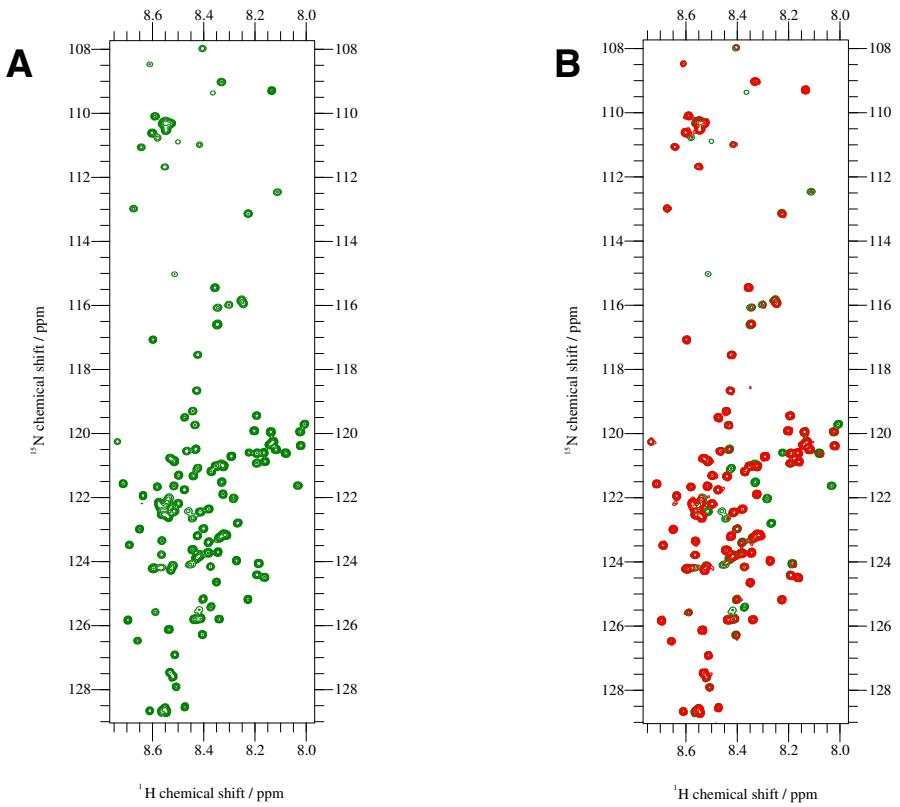

Figure 5.1: ^{1}H-^{15}N *HSQC spectra of* H_6-α*syn.* **A:** Isolated α-synuclein. **B:** Isolated α-syn (green) at 5 overlaid with a spectrum of α-syn in the presence of a 1:1 mixture (5 μM:5μM) of ribosomes (red). Spectra were acquired at 700 MHz, pH 7.5 and 4 °C.

in small amounts might affect the interaction between α-syn and the ribosome. Cells were lysed by French press, and the lysate was purifed by ultracentrifugation through a 30 % sucrose cushion followed a 10-30 % sucrose gradient. Sample purity was assessed by SDS-PAGE.

5.1.2 α-Synuclein line-broadening in the presence of ribosomes

α-Synuclein gives a good quality spectrum, with a narrow amide proton region, diagnostic of an intrinsically disordered protein (Figure 5.1, left). On mixing with ribosomes, HSQC cross-peaks appear to undergo substantial line-broadening effects associated with a sharp increase in transverse relaxation rate, R_2 (Figure 5.1, right). Comparison of the intensity of HSQC peaks in the presence of ribosomes relative to that in isolation shows that line broadening is most significant at the C and N-termini (Figure 5.2).

To evaluate the interaction in detail, titrations were carried out on both wild-type and his-tagged synuclein constructs against ribosomes, and binding was assessed by measurement of amide proton relaxation rates

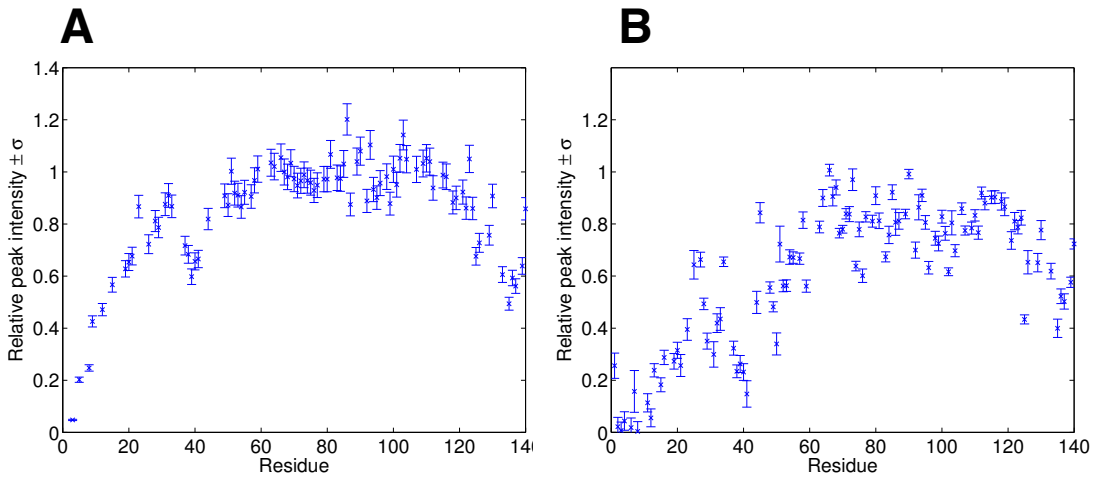

Figure 5.2: *α-syn HSQC intensity plots in presence of ribosomes.* Relative intensities of HSQC cross-peaks for WT (**A:**) and his-tagged (**B:**) α-synuclein in the presence of ribosomes at a 1:1 ratio. A relative intensity of 1 indicates no broadening in the presence of ribosomes. The line-broadening can be seen to be far more effective in the presence of the histidine tag. Spectra were collected at 700MHz, 4 °C and pH 7.5 in Tico buffer.

using a spin-echo ^{1}H-^{15}N HSQC method, as described by Iwahara et al. [122]. The observed relaxation rates obtained for each residue were fit globally to a 1:1 model, described by equation 5.1.1 (for details on the theory and data fitting, see section 5.2),

$$R_{2,obs} = \frac{1}{2}\left[k_{ex} + 2R_{2A} + \Delta R_{2,b} - \sqrt{k_{ex}^2 + 2k_{ex}\Delta R_{2,b}(1-2p_B) + \Delta R_{2,b}^2}\right] \tag{5.1.1}$$

where $R_{2,obs}$ is the observed relaxation rate, R_{2A} is the relaxation rate of isolated synuclein, $\Delta R_{2,b}$ is the change in relaxation rate on binding ribosomes ($= R_{2B} - R_{2A}$), and k_{ex} is the rate of exchange between bound and unbound species.

Each residue was treated individually to give a best fit in terms of $R_{2,b}$, the dissociation constant K_d and the association constant k_{on}, using a χ^2 minimisation procedure. Global fits are shown in Figure 5.3. This revealed an optimal value of $\Delta R_{2,b}$ for each residue, as well as global values for the K_d and k_{on}. For wild-type α-synuclein, this suggests a minimum association rate k_{on} of 10^6 M^{-1} s^{-1} while the K_d appeared to be in the range of 1-20 mM. Repeating the same analysis for H_6-αsyn gives a range for the K_d between 300 μM and 16 mM, while association kinetics remained the same.

Although no upper bound was observed for kinetics, one can estimate the maximum association rate to be 10^8 M^{-1} s^{-1} as this is an estimated maximum diffusion limit arising from consideration of Smoluchowski

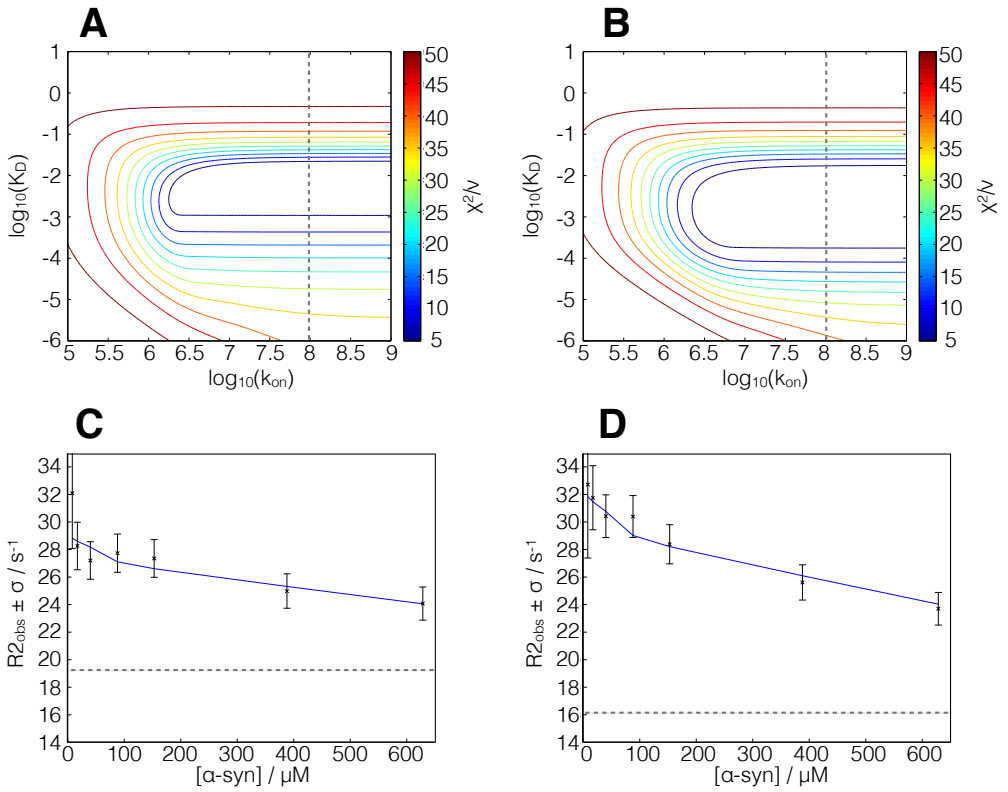

Figure 5.3: *Global χ^2 surfaces and fits for R_2 as a function of concentration.* **A:** Global χ^2 surface obtained from spin-echo relaxation experiments for wild-type α-synuclein. Dotted lines represent the estimated diffusion limited association rate. **B:** Global χ^2 surface obtained from spin-echo relaxation experiments for wild-type α-synuclein. Dotted lines represent the estimated diffusion limited association rate. Areas with χ^2 lower than 30 (green contours) represent good quality fits to the binding model, and so it is clear that for the K_d, a very broad range of values is allowed. **C:** Fitted curve showing the agreement between calculated and observed relaxation rates (equation 5.1.1) for residue 38. **D:** Fitted curve showing the agreement between calculated and observed relaxation rates (equation 5.1.1) for residue 135. Grey dotted lines indicate the R_2 of free synuclein.

diffusion modified to include electrostatic interactions [123]. The area in which the K_d and k_{on} must lie is therefore constrained for both constructs (see Figure 5.3).

With these association rate limits, the overall exchange rate k_{ex} can be shown to be between 5.2×10^3 and 1.2×10^4 s^{-1}. This relatively high exchange rate, combined with the somewhat weak interaction, explains why it is still possible to observe significant line-broadening at 800:1 molar excesses of H_6-αsyn, since many molecules of synuclein are able to transiently bind to a single ribosome in a small time window, leading to a build up of line-broadening effects.

These results indicate the presence of an extremely weak interaction with diffusion limited association kinetics. Although the interaction may seem too weak to be biologically relevant, the local concentration of

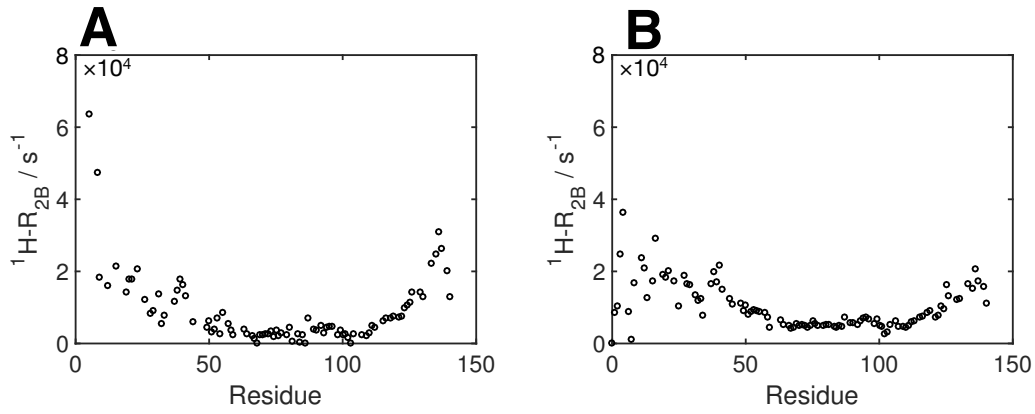

Figure 5.4: *Ribosome-bound state relaxation rates of α-syn.* Plots of the bound state R_2, ($R_{2,b}$) of (**A**) wild type and (**B**) his-tagged α-synuclein in the presence of ribosomes. $R_{2,b}$ values were extracted from the data fitting procedure described in section 5.3.

a nascent chain undergoing translation at the ribosomal surface is likely to be in the millimolar range, and hence the interaction may still have the potential to play a role in affecting the folding pathway of translating proteins. It is also clear that the polyhistidine tag has a notable effect on the interaction, which indicates that the routine use of this sequence for purification should be considered with some caution. It is likely that the highly positive charge of the his-tag enhances the affinity for the ribosomal surface, which is largely negatively charged. However, other his-tagged constructs, such as FLN750, do not show such a substantial change in peak intensity in the presence of ribosomes, suggesting that electrostatics is not the only factor one should consider when investigating these interactions.

Inspection of the bound state relaxation rates determined from this analysis (see Figure 5.4) strongly indicates that binding occurs at the N-terminus, consistent with the reduction in peak intensities. Note that for the first twenty residues the relaxation rate $\Delta R_{2,b}$ exceeds or is close to the estimated k_{ex}. This indicates an intermediate to slow exchange regime at the N-terminus, verifying the use of the Bloch-McConnell equations for calculating these relaxation rates. Beyond the 20^{th} amino acid, fast exchange kinetics begins to dominate.

To conclude, it has been shown that global fitting of residue specific intensity data to a 1:1 interaction model can successfully describe the interaction between H_6-αsyn and the ribosome, and this also leads to a successful determination of the bound state relaxation rate. Although we cannot confirm the proposed binding model at this stage, the fact remains that this is the simplest binding model that fits to the current data, and so no further complexity, if it exists, can be investigated at this stage.

Although it is not clear what role this interaction might play in co-translational folding, the work described here is an important step towards uncovering the nature of protein-ribosome interactions, and shows that they can be highly dynamic, transient processes as opposed to static events that fix a protein in a specific conformation for a long duration of time. Perhaps a protein only needs a small amount of time at the ribosome surface for there to be a significant impact on the overall conformational landscape. If the ribosome surface presents an environment that is markedly different to that found in the cell cytoplasm, then only a small amount of time might be required for substantial changes in conformation to be seen. However, the bound state was not directly observed here, and so it is difficult to evaluate whether it is notably different from the free state. We can only be sure that it is less dynamic, due to the high transverse relaxation rates that were measured for the bound state.

As of the moment, there is no means of directly probing, by NMR, the nature of the ribosome-bound state of a protein, since relaxation is simply too fast. However, in section 5.4 we describe the use of fluorine NMR to investigate co-translational folding, and show how this unique nucleus may be able to give direct chemical shifts of ribosome-bound species.

5.2 Background theory for analysis of α-syn titration data

By solving the Bloch-McConnell equations to account for kinetics, we show that the α-syn ribosome interaction (Section 5.1) is represented well by fitting to a global K_d and k_{ex} alongside residue specific relaxation rates $\Delta R^i_{2,b}$. Acknowledgements to Dr John Kirkpatrick for help with the background theory and data analysis provided in this section.

5.2.1 Solution of Bloch-McConnell equations for weak binding to a large species

Given the mass of the ribosome relative to α-syn, the change in correlation time of α-syn residues upon binding is great enough to induce relaxation rates comparable to or greater than the exchange rate between the free and bound protein, k_{ex}.

In order to account for this intermediate/slow exchange regime, we must seek an appropriate solution to the Bloch-McConnell equations, taking into account the regular parameters of the Bloch equations (Larmor frequency, transverse relaxation) and including also an exchange term to account for kinetics.

To appropriately include kinetics, we must also relate exchange rates to equilibrium parameters. Considering the synuclein-ribosome interaction (equation 5.2.1) from the perspective of the protein alone, we define

two states, A and B, corresponding to free and bound protein, respectively (equation 5.2.2).

$$P + R \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} PR \tag{5.2.1}$$

$$A \underset{k_{BA}}{\overset{k_{AB}}{\rightleftharpoons}} B \tag{5.2.2}$$

The overall exchange rate k_{ex} can be described as the sum of the pseudo first order rate constants k_{AB} and k_{BA},

$$k_{ex} = k_{AB} + k_{BA} \tag{5.2.3}$$

and the relation between kinetics and equilibrium is;

$$K_d = \frac{k_{off}}{k_{on}} \tag{5.2.4}$$

The off rate, k_{off}, is equivalent to k_{BA}, as both are concentration independent. The forwards rate k_{AB}, however, is dependent on the concentration of free ribosomes at equilibrium, $[R]_{eq}$,

$$k_{AB} = k_{on}[R]_{eq} = k_{on}\left([R]_0 - [PR]\right) \tag{5.2.5}$$

where [PR] is given by;

$$[PR] = \left(\frac{[P]_0 + [R]_0 + K_d - \sqrt{([P_0] + [R]_0 + K_d)^2 - 4[P]_0[R]_0}}{2}\right). \tag{5.2.6}$$

These considerations allow us to build an exchange model based on observing only H_6-αsyn (equation 5.2.2), from which we can determine the parameters of the whole interaction (equation 5.2.1).

For two precessing spins A and B undergoing exchange, the full Bloch-McConnell matrix, accounting for contributions from chemical exchange (k_{ex}), transverse relaxation ($R_{2A/B}$) and difference in larmor frequency ($\Delta\omega$), can be written thus;

$$\frac{d}{dt}\begin{pmatrix} M_A^+ \\ M_B^+ \end{pmatrix} = \begin{pmatrix} -R_{2A} - k_{AB} & k_{BA} \\ k_{AB} & i\Delta\omega - R_{2A} - \Delta R_{2,b} - k_{BA} \end{pmatrix} \cdot \begin{pmatrix} M_A^+ \\ M_B^+ \end{pmatrix} \tag{5.2.7}$$

We can denote the original $\begin{pmatrix} M_A^+ \\ M_B^+ \end{pmatrix}$ matrix as $\overrightarrow{\mathbf{M}}(t)$, and the relaxation/exchange matrix as $\mathbf{A}$ in which case

the expression becomes;

$$\frac{d\overrightarrow{\mathbf{M}}(t)}{dt} = \mathbf{A} \cdot \overrightarrow{\mathbf{M}}(t) \tag{5.2.8}$$

which is solved by taking the matrix exponential of $\overrightarrow{\mathbf{M}}$,

$$\overrightarrow{\mathbf{M}} = \mathrm{e}^{\mathbf{A}t}\overrightarrow{\mathbf{M}}(0) \tag{5.2.9}$$

Where $\overrightarrow{\mathbf{M}}(0) = \left(\begin{smallmatrix}\mathrm{p_A}\\ \mathrm{p_B}\end{smallmatrix}\right)$, and $\mathrm{p_A}$ and $\mathrm{p_B}$ represent molar fractions of A and B, respectively.

The assumption that $\mathrm{p_B} \ll 1$ allows the observed spectrum to be described purely in terms of the A state magnetisation, which we express as a linear combination of two exponentials, λ_1 and λ_2, the eigenvalues of $\mathbf{A}$. It can be shown that;

$$\lambda_{1,2} = \frac{1}{2}\left[-k_{ex} - 2R_{2A} - \Delta R_{2,b} + i\Delta\omega \mp \sqrt{k_{ex}^2 + 2k_{ex}(1-2p_B)(\Delta R_{2,b} - i\Delta\omega) + (\Delta R_{2,b} - i\Delta\omega)^2}\right] \tag{5.2.10}$$

We can simplify this model, however, by neglecting chemical exchange, assuming $\Delta\omega \ll \Delta\mathrm{R_{2,b}}$. The eigenvalues then become;

$$\lambda_{1,2} = \frac{1}{2}\left[-k_{ex} - 2R_{2A} - \Delta R_{2,b} \mp \sqrt{k_{ex}^2 + 2k_{ex}\Delta R_{2,b}(1-2p_B) + \Delta R_{2,b}^2}\right] \tag{5.2.11}$$

Recalling the assumption $\mathrm{p_B} \ll 1$ allows us to obtain a first order Taylor expansion in $\mathrm{p_B}$;

$$\Delta R_{2,obs} = \frac{k_{ex}\Delta R_{2,b}}{k_{ex} + \Delta R_{2,b}} \cdot p_B \tag{5.2.12}$$

(Given that $\lambda_1 \equiv -\mathrm{R_{2,obs}}$). This simplified expression allows us to note the presence of two limits, according to the relative sizes of $\Delta\mathrm{R_{2,b}}$ and $\mathrm{k_{ex}}$,

$$k_{ex} \gg \Delta R_{2,b} \implies \Delta R_{2,obs} \to \Delta R_{2,b} p_B \tag{5.2.13}$$

$$k_{ex} \ll \Delta R_{2,b} \implies \Delta R_{2,obs} \to k_{ex} p_B \tag{5.2.14}$$

We note the unusual nature of this slow exchange regime, in that it is due to $\mathrm{k_{ex}} \ll \Delta\mathrm{R_{2,b}}$ as opposed to the more common $\mathrm{k_{ex}} \ll \Delta\omega$. This particular kind of extreme is often observed in the case of paramagnetic relaxation [78], but besides this it is rare to observe slow exchange regimes caused by high transverse relaxation rates.

5.3 Minimisation of binding parameters by χ^2 analysis

For a given residue i, the χ^2 value between intensity that is measured and that which is calculated from trial binding parameters K_d, k_{ex} and $R^i_{2,b}$ is given by

$$\chi^2_{\text{min,i}} = \sum_{\text{titration points}} \left(\frac{[R_{2,obs}]_{\text{calc}} - [R_{2,obs}]_{\text{obs}}}{\sigma} \right)^2 \tag{5.3.1}$$

and this can be minimised for all parameters for each individual residue i. We then define the global χ^2 as,

$$\chi^2_{\text{global}} = \sum_{\text{resi}} \chi^2_{\text{min,i}}\{K_d, K_{on}\} \tag{5.3.2}$$

excluding $\Delta R_{2,b}$ since this is determined from $\chi^2_{\text{min,i}}$ and then fixed to a set value, allowing us to treat the global fit as a sum of χ^2 fits for each amino acid. The principal eigenvalue (λ_1) from equation 5.2.11 was used to calculate trial relaxation rates ($R_{2,b}$) from test values of the binding parameters, by expressing λ_1 as a function of K_d, k_{on} and $R_{2,b}$.

5.4 Advantages and applications of the fluorine nucleus

The use of ^{19}F as a label in NMR spectroscopy has a wide and varied history, owing to its high natural abundance, large gyromagnetic ratio (94 % that of the proton), and the high sensitivity of its chemical shift to changes in the local environment. The nucleus has been utilised for NMR based studies of protein-protein and protein-ligand interactions, protein folding, and drug screening techniques. Since biology is almost devoid of fluorine, the selective incorporation of this element into proteins allows NMR spectra to be collected on very specific chemical moieties without any background interference from biological systems.

The fluorine nucleus has been used in a broad range of applications. It has been utilised as a basic probe of macromolecular structure [124] and protein structure and conformation [125, 126], as well as a means of studying protein-protein interactions [127, 128], and protein folding [129, 130]. Work is also being developed to exploit this nucleus for in-cell studies [131–134], exploiting the advantage that fluorine spectra have very little background signal.

The most common form of fluorine incorporation for protein NMR is to use it to replace a proton. This has little effect upon sterics, since the atoms have very similar Van der Waals radii. This similarity allows the above advantages for fluorine to be exploited in biological systems, while also treating the fluorine modified

systems as similar to their non-fluorinated counterparts. However, the C-F bond is notably more polar than a typical C-H bond, and so some chemical differences between the two analogs may occur.

Incorporation of fluorine is also used for pharmacological purposes, where the fluorine moiety is intentionally exploited due to its unique chemical properties [135]. This points to a potential issue with the use of fluorine as an NMR probe, since it is not identical to the proton in terms of chemical behaviour, and can affect protein structure. In extreme cases, it has been used to drastically alter the fold and properties of a protein such that it is entirely different from the wild type [136]. As such, care must be taken to ensure that structural and folding information acquired from fluorine labelled samples actually pertains to a biologically relevant state of the protein.

The major disadvantage of fluorine lies in its high chemical shift anisotropy (CSA) [129], which leads to extremely broad resonances for molecules with long correlation times or at higher magnetic fields, and also makes certain types of 2D experiments extremely difficult to collect [137]. Therefore, in the context of using fluorine to study nascent chains there is a concern that this will cause resonances on the ribosome to become unobservably broad when the high CSA is combined with the long correlation time of the massive ribosome. However, even very broad resonances could yield information about the nascent chain, provided they are visible at all. This is related to another principle advantage of fluorine, in that it can yield very clean NMR spectra with only a few peaks, which only correspond to the desired protein. Therefore, provided we know the origin of a very broad resonance, the absence of interfering peaks would allow us to analyse it in a way that would be very difficult in proton NMR, for example, where interference from background and from solvent could render very broad resonances impossible to observe, let alone interpret.

5.5 Methods for incorporation of fluorine into protein samples

In order to incorporate fluorine into an expressed protein, one typically uses a strain of bacteria that has been rendered auxotrophic for the amino acid one wishes to substitute. A common means of inducing auxotrophy in otherwise healthy *E. coli* strains is to include N-(phosphonomethyl)glycine (or glyphosate) in the growth medium. This broad spectrum herbicide inhibits the enzyme 5-enolpyruvylshikimate-3-phosphate synthase, which is crucial for the synthesis of aromatic amino acids [126]. Thus, by then ensuring the only source of aromatics is fluorine substituted, expression can then be induced in the conventional manner, leading to high levels of fluorine incorporation. Although it does not give the highest substitution yields, this approach is a simple and easy means of obtaining incorporation of non-canonical analogues of the three aromatic amino

acids.

An alternative approach is the selective pressure incorporation (SPI) method, where genetically engineered bacterial strains that are auxotrophic for a specific amino acid are incubated in the presence of the desired non-canonical amino acid. Provided that the new amino acid bears enough similarity to the original, it can be incorporated into the same tRNA. This approach has been successfully utilised for the incorporation of numerous amino acid analogs, including trifluoromethionine [138, 139], numerous proline derivatives [140], fluorohistidine [124], difluoromethionine [125], 4-fluorotryptophan [141], and others.

5.6 Fluorine NMR on isolated proteins

Prior to the production of fluorinated nascent chains, initial investigations into the fluorine NMR properties of isolated proteins were carried out. In particular, we chose to investigate α-synuclein and domain 5 of the gelation factor (FLN750). In order to incorporate fluorinated aromatic amino acids into the protein, BL21 cells transformed with the required expression plasmid were grown in a minimal M9 medium (with the desired labelling, or unlabelled) and when the optical density reached 0.45, the medium was supplemented with glyphosate and a combination of phenylalanine, tryptophan, and 3-fluorotyrosine. This method, adopted from Li et al. [126], was used to produce samples of α-synuclein and ddFLN750 substituted with 3-fluorotyrosine. See section 7.1.7 and 7.1.8 for experimental details. NMR experiments were carried out on a Bruker Avance 600 spectrometer equipped with a fluorine probe. All chemical shifts were externally referenced to trifluoroacetic acid (TFA) at -76.55 ppm.

5.6.1 Fluorine NMR of α-synuclein

1D fluorine spectra of ^{19}F-^{15}N labelled α-synuclein show the presence of three adjacent peaks at -136.75, -136.86 and -136.96 ppm (see Figure 5.5). The peaks appear identical to those identified in the literature for the same construct [126], after accounting for chemical shift referencing. Hence, while there are four fluorotyrosine residues in α-syn, we only expect to see three peaks under the conditions used here. The disordered nature of α-syn causes the peaks to cluster at similar chemical shifts. It is possible to resolve the fourth peak by mixing the α-synsample with SDS-micelles [126], however the agreement between the spectra observed here and in the literature was sufficient to confirm a good quality sample.

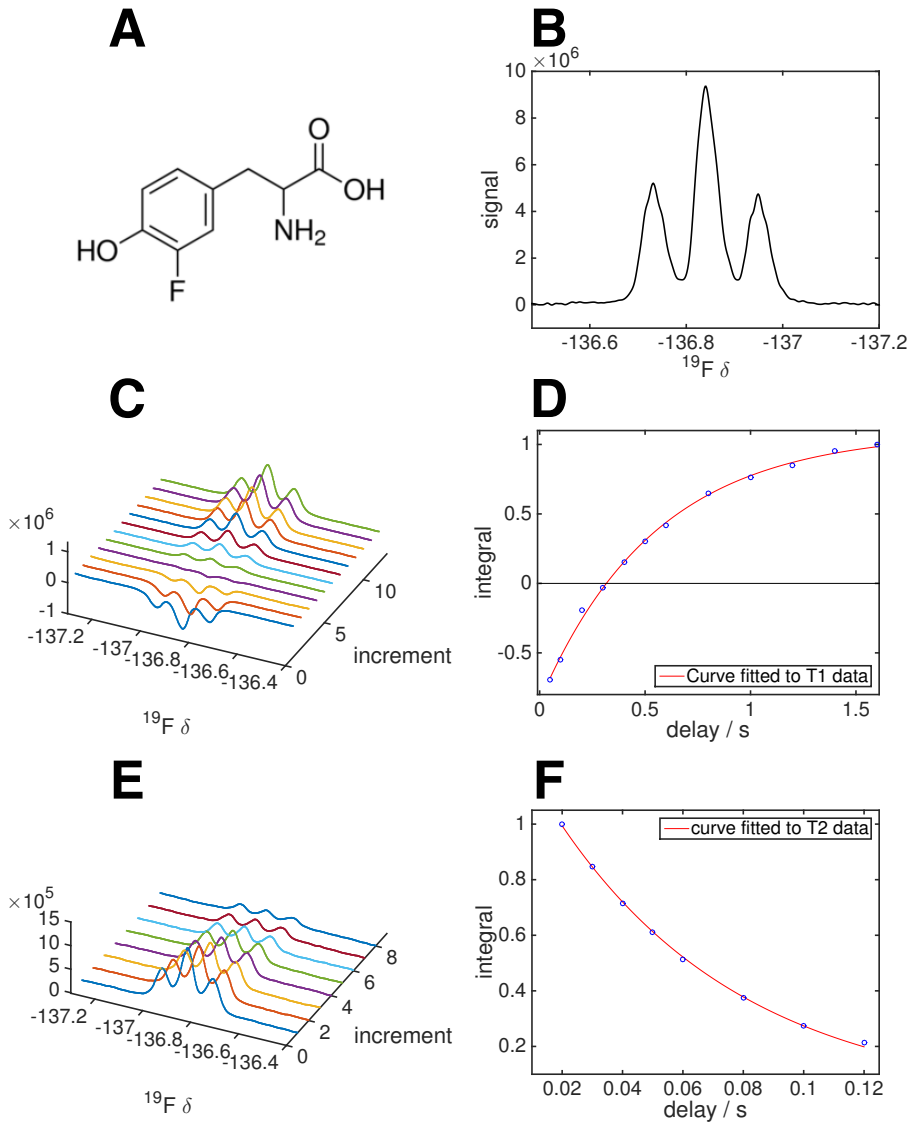

Figure 5.5: ^{19}F *labelled spectra of 3-fluorotyrosine labelled* α*-synuclein.* **A:** Chemical structure of 3-fluorotyrosine. **B:** 1D ^{19}F spectrum of α-syn. **C:** T_1 inversion recovery experiment. **D:** T_1 inversion recovery fit. **E:** T_2 CPMG experiment. **F:** T_2 CPMG fit. All spectra were acquired at 600 MHz, pH 7.5 and 25 °C.

5.6.2 Fluorine NMR of FLN750

^{19}F NMR of FLN750 were also investigated and subjected to similar analyses as α-syn. Four peaks of similar linewidth were identified, two of which are strongly overlapped (Figure 5.6). The peaks appear at -135.78, -136.30, -13.44 and -135.2 ppm. There was also a fifth, much broader peak centred at -135.2 ppm, the identity of which is uncertain. The presence of five peaks is also not consistent with the presence of only 4 tyrosine residues in FLN750.

To investigate the impact of unfolding on the fluorine chemical shift, a 3-fluorotyrosine labelled variant of FLN750 with a 12 amino acid C-terminal truncation was produced, termed FLN738. This protein is known to be completely unfolded at room temperature. Inspection of ^{19}F 1D spectra showed signal characteristic of an unfolded protein, as the fluorine signals have collapsed into a far narrower range of peaks, with only minimal dispersion (Figure 5.6).

5.6.3 Relaxation measurements of the ^{19}F signal on FLN750 and α-syn

Samples of FLN750 and α-syn were investigated with T_1 inversion recovery and T_2 CPMG relaxation measurements. At 25 °C, the fluorine signal from α-syn yielded a T_1 of 548.9 ± 59.1 ms, and a T_2 of 62.1 ± 2.2 ms. Fluorine signal from FLN750 gave a T_1 of 597.4 ± 59.7 ms and a T_2 of 6.9 ± 1.0 ms.

The T_1 values are similar for both proteins since they are of similar size. However, the T_2 of α-syn is 1 order of magnitude slower than FLN750, as the more dynamic, disordered α-syn has a shorter correlation time than the rigid, structured domain 5.

Due to the high CSA of the fluorine nucleus, one can expect the T_2 to drop very sharply at higher magnetic fields, as well as for larger molecular weight species. As such, in the case of nascent chain studies, it will be important to investigate such samples using weaker fields than those typically used for nascent chains, such as 500-600 MHz, and to minimise CSA effects by choosing to study systems where dynamic behaviour (such as is found in a disordered, or unfolded protein) can reduce the local correlation time of the nascent chain. Therefore, it may be the case that ^{19}F is better suited for the investigation of unfolded nascent chains, such as α-syn or the FLN750 nascent chains at shorter linker lengths where the unfolded state is populated on the ribosome.

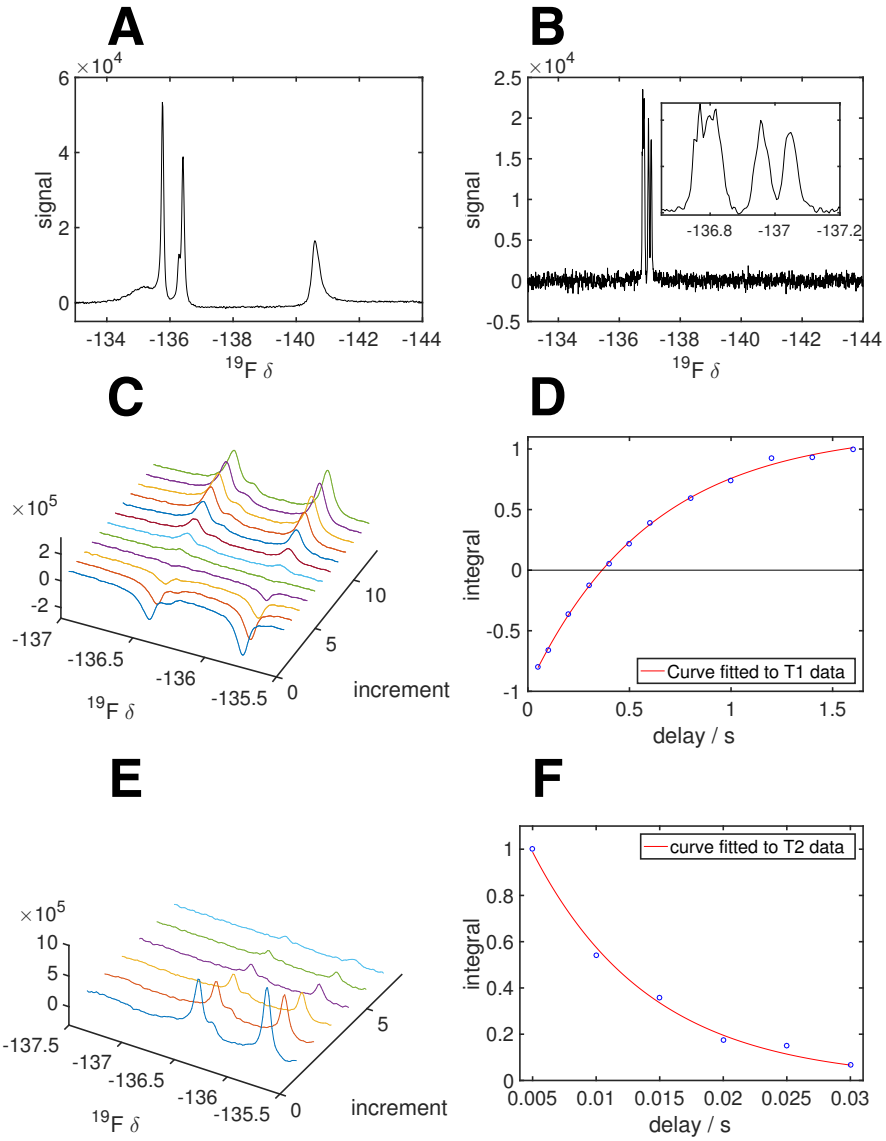

Figure 5.6: *^{19}F labelled spectra of 3-fluorotyrosine labelled domain 5.* **A:** 1D ^{19}F spectrum of domain 5. **B:** 1D ^{19}F spectrum of the truncated form of domain 5 (FLN738), and inset showing the fine structure of the peaks at -137 ppm. **C:** T_1 inversion recovery experiment of FLN750. **D:** T_1 inversion recovery fit. **E:** T_2 CPMG experiment of FLN750. **F:** T_2 CPMG fit. All spectra were acquired at 600 MHz, pH 7.5 and 25 °C.

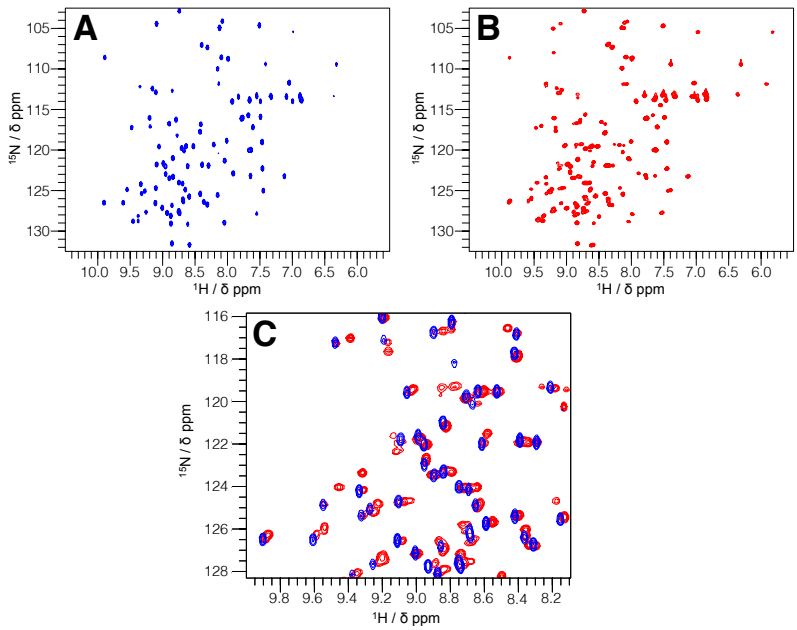

Figure 5.7: *^{1}H-^{15}N NMR spectra of the ^{19}F labelled FLN750 protein.* **A:** ^{1}H-^{15}N HSQC spectrum of the ^{15}N labelled FLN750 domain. **B:** ^{1}H-^{15}N HSQC spectrum of the ^{19}F-^{15}N labelled FLN750 domain. **C:** Overlay of the two spectra (same colour code), zoomed in to highlight the presence of extra, shifted peaks in the spectrum of the ^{19}F labelled construct.

5.6.4 ^{1}H-^{15}N NMR spectra of the FLN750 protein

For further investigation of the FLN750 protein prior to production of the nascent chain, ^{19}F-^{15}N labelled samples were expressed and purified. HSQC spectra of the FLN750 construct show a good overlay with the non-^{19}F labelled variant, indicating the presence of the structured protein (Figure 5.7). However, also observed were numerous extra peaks, and doubled peaks, which is typically indicative of multiple species. This may be due to differential fluorine labelling, whereby the presence of numerous species leads to multiple environments for peaks in proximity to the fluorine substitution sites.

The ^{19}F and ^{1}H-^{15}N NMR spectra of FLN750 are not immediately indicative of the nature of the sample produced. Considering the ^{19}F 1D spectra, the presence of five peaks when there are only four tyrosines in the primary sequence could be explained by some form of conformational exchange induced by rotation of one of the fluorotyrosine side-chains, with exchange broadening effects leading to the highly increased line-width. It is also possible that some scrambling of the labelling occurred, which would lead to fluorine at other positions. However, one would expect this to lead to many more than five peaks in the fluorine spectrum, and this scrambling was not observed for ^{19}F labelled α-syn, which was produced in the same way.

Considering also the ^{1}H-^{15}N HSQC spectra (Figure 5.7), it is clear that the fluorinated domain 5 sample has many more peaks in the ^{15}N HSQC. The origin of these peaks is not clear. It may be the case that differential degrees of fluorination - where some molecules have more fluorine substituted sites than others - has lead to multiple species, and for each species there is therefore a slightly different chemical environment presented to the remaining nuclei. This might cause a single peak to become shifted depending on whether the nuclei of that peak are proximal to tyrosine or a fluorotyrosine. This could be probed with a more precise fluorination approach, where only one amino acid is substituted for its fluorinated analog, or by mutational knockouts of selected tyrosines, allowing investigation of the impact of fluorination on the remaining tyrosines and other proximal residues.

Alternately, it may be the case that fluorination of domain 5 leads to some destabilisation of the protein, causing conformational exchange with other states. This would lead to multiple environments for each nucleus as a result of the exchange process. This could be investigated further with NMR experiments such as relaxation dispersion [130] or saturation transfer experiments, which are able to probe conformational exchange processes.

5.7 Producing a fluorine labelled nascent chain

The α-synuclein and FLN750 nascent chains have both been very well developed in terms of expression and purification approaches, and so either protein stands as a strong candidate for exploiting the fluorine nucleus. The FLN750 nascent chain was chosen in the first instance because it allows us to investigate the advantages of fluorine both in a folded and unfolded context, depending upon linker length. Therefore, the +31 and +110 FLN750 linker lengths were chosen for initial ^{19}F NMR studies as these represent wholly structured and wholly unstructured translationally stalled nascent chains.

5.7.1 Method for the effective production of fluorine labelled nascent chains

To incorporate fluorinated amino acid analogs into RNC samples, the standard protocol for growing RNCs was adapted by inclusion of a step involving glyphosate; cells transformed with the intended nascent chain construct were grown in unlabelled MDG medium overnight. After 16-18 hours, the cells were harvested ($OD_{600} = 4 - 6$) and resuspended in EM9 medium. Isotopes were added last, alongside the glyphosate and aromatic amino acids. The cells were incubated for 15 minutes to ensure that the OD was still increasing, and IPTG was added to induce expression. After ten minutes of expression, rifampicin was added, followed

by a further 35 minutes of incubation, and the cells were harvested. Growth and expression was all carried out at 37 °C.

Purification of the nascent chains was carried out as described in Methods section 7.5.3. Briefly, this involved a rapid sucrose cushion, followed by nickel affinity chromatography, a second cushion, and final purification on a butyl column. The samples were then flash frozen in liquid nitrogen and stored at -80 °C.

5.7.2 Labelled ribosome growth

Positive controls of fluorinated ribosomes were also grown. Ribosome fluorination was carried out via two approaches; to maximise the possibility of fluorination, untransformed *E. coli* were grown to an OD_{600} of 0.6 in MDG before supplementation with glyphosate and amino acid analogs, and growing the culture to saturation over 16 hours. Cells were then harvested and purified as normal. Alternately, untransformed cells were treated to an identical procedure as given for the nascent chain growth; grow to saturation in MDG, transfer to EM9 and induce, addition of of supplements at identical time points etc. Both these methods act as controls to assess whether fluorination of ribosomes leads to observable signal that could interfere with signal from the nascent chain.

5.8 ^{15}N NMR of the ^{15}N-^{19}F labelled FLN750 +31 RNC.

5.8.1 Sample preparation

The frozen sample was thawed on ice, and buffer exchanged into fresh Tico buffer (5 ×, 0.5 - 15 ml) to filter out nascent chain that can release in the freeze/thawing process. The final NMR sample (including 10 % D_2O and 0.01 % DSS) contained 9.8 µM ribosomes with a volume of 300 µl.

5.8.2 NMR spectra of the RNC

All fluorine spectra were recorded at 10 °C on a 600 MHz Bruker Avance spectrometer equipped with a cryogenic probe and a fluorine sensitive channel. SOFAST HMQC spectra were recorded with 1024 time points in the direct dimension and 64 points in the indirect (^{15}N) dimension. The centre of the spectrum in the indirect dimension was set to 117 ppm with a sweep width of 22 ppm, corresponding to a resolution of 0.25 ppm (35.5 Hz).

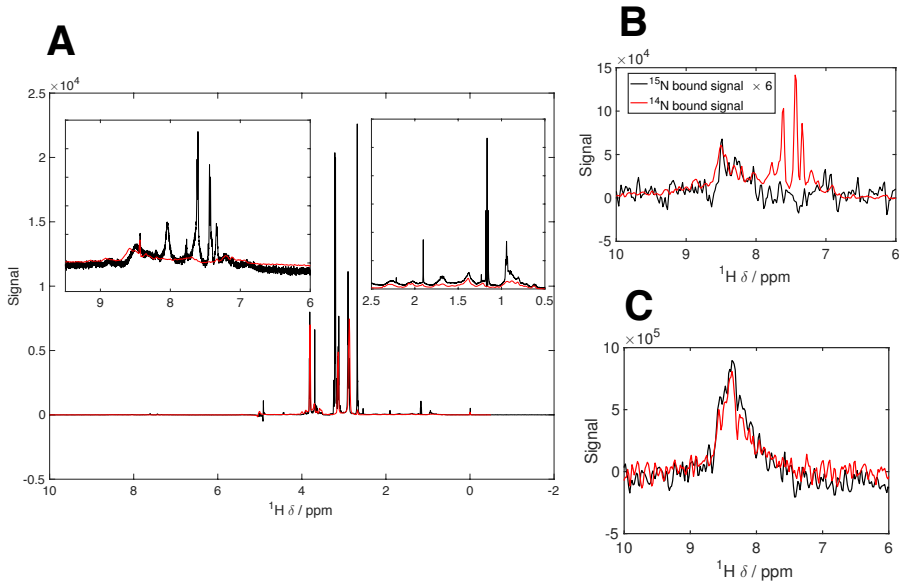

Figure 5.8: *Initial 1D ^1H spectra of the ^{19}F labelled nascent chain.* **A**: 1D ^{1}H spectrum of the ^{19}F nascent chain (black) overlaid with 1D ^{1}H spectrum of unlabelled 70S ribosomes. The similarities in patterns between the spectra indicates the presence of ribosomes in the ^{19}F labelled sample. **B**: ^{15}N edited (black) and filtered (red) HMQC 1D spectra of the nascent chain, scaled to equal intensities for evaluation of the extent of background labelling. **C** ^{1}H-^{15}N SORDID-diffusion experiments on ^{19}F labelled nascent chain with gradient strength G at 100 % G_{max} (red) and 15 % G_{max} (black).

5.8.3 Initial evaluation of sample quality and integrity

Prior to acquisition of fluorine spectra, the sample was evaluated by standard ^{1}H and ^{15}N spectra in order to evaluate the quality of the sample via a means that is comparable to other nascent chain experiments. The sample was therefore evaluated using 1D ^{1}H, 1D ^{15}N edited and filtered SOFAST spectra, and ^{15}N edited SORDID diffusion experiments. In other words, it was treated to the standard ^{15}N analysis of a nascent chain (see Section 4.5 for the GA module experiments). These spectra are summarised in Figure 5.8.

Inspection of the 1D ^{1}H spectrum reveals a profile characteristic of purified ribosomes, as shown by overlaying the RNC 1D spectrum with a 1D spectrum of previously purified ribosomes (Figure 5.8A). There are also additional peaks characteristic for the nascent chain, that have been observed in previous samples. There were also some impurities present in the nascent chain sample that could not be identified. However, the sample was still deemed as sufficient in quality for further analysis by NMR.

The dual experiments of ^{15}N filtered and ^{15}N edited 1D spectra were carried out to highlight the extent

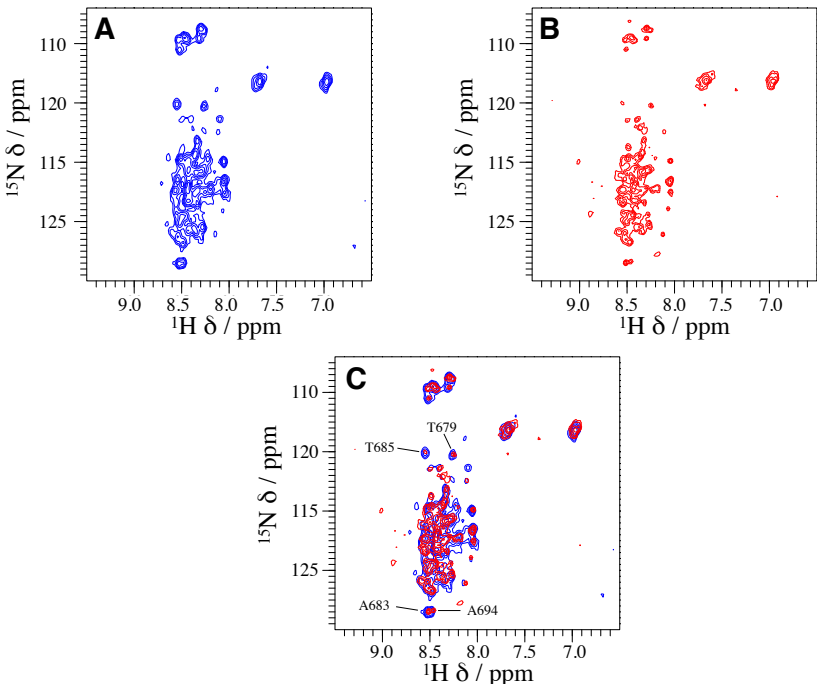

Figure 5.9: *^{1}H-^{15}N HMQC spectra of the ^{19}F-^{15}N labelled +31 domain 5 (FLN750) nascent chain.* **A:** Spectrum of the ^{15}N labelled nascent chain, with no fluorine substitution. This acts as a control spectrum for fluorination. Spectral data was provided by Anaïs Cassaignau. **B:** Spectrum of the fully ^{19}F-^{15}N labelled nascent chain. **C:** Overlay of both spectra, highlighting the high degree of agreement between the two samples. Note the appearance in both spectra of peaks corresponding to unfolded domain 5 residues.

of ^{15}N labelling of the L7/L12 stalk region of the ribosome. From these spectra, it was estimated that the background labelling present in the ribosome was 16 % of that of the nascent chain.

SORDID diffusion allowed measurement of the diffusion coefficient D arising from the amide proton signal, which hence corresponds to the diffusion of the labelled nascent chain. The value of D obtained from fitting to the Stejskal-Tanner equation (see Equation 4.5.1, Section 4.5.2) was 2.5×10^{-12} m^2 s^{-1}. This is far smaller than expected for a nascent chain, indicating a particularly large species, however this could be attributed to error associated with the diffusion due to weak signal. Finally, SOFAST HMQC spectra [61] were also acquired to compare observed resonances with those from previous samples (Figure 5.9).

5.9 Fluorine NMR of the nascent chain

After the initial confirmation of sample quality with the ^{1}H-^{15}N control experiments, it was possible to acquire fluorine NMR spectra with confidence about the identity of the sample being investigated.

The sample was analysed by the collection of ^{19}F 1D, T_1 relaxation inversion recovery and DOSY diffusion experiments. The spectra were recorded continuously over a period of 15 hours in an interleaved manner in order to investigate how the sample was changing over time. Each round of 1D, diffusion and relaxation experiments took 1.2 hours to collect.

The 1D ^{19}F spectra show a strong signal at -119.2 ppm, which was observed to increase in intensity over time as the experiment was repeated (Figure 5.11A-B). This signal fits well to a Lorentzian lineshape with a linewidth of 121 ($\pm$1.86) Hz for the first spectrum that was collected, 1 hour after the sample was thawed and buffer exchanged (Figure 5.11C). The linewidth of the signal was also observed to increase over time along with the peak intensity (Figure 5.11D).

T_1 inversion recovery relaxation experiments were also acquired to determine whether this property could yield meaningful information on sample quality over time (Figure 5.10). These relaxation experiments indicated a T_1 value of around 1400 ms. However, the error in the measurement of the T_1 is too large to determine how it changes over time. DOSY diffusion experiments were also collected, however they gave such weak signal that no information on the diffusion of the species could be obtained.

5.9.1 Monitoring sample integrity over time

Alongside the NMR experiments, the sample was also monitored by a parallel western blot analysis (Figure 5.11E-H). As the sample was prepared for NMR, a small aliquot of 20 pmol was taken and kept under exactly the same conditions as the main sample. This was divided into further 3.5 pmol aliquots which were frozen at the same time each SORDID experiment was collected, allowing for parallel anti-his western blots to be carried out at a later date (Figure 5.11E). Densitometry analysis of the western blot was performed using the imageJ software, which revealed the intensity of the tRNA bound form to remain relatively constant over time (Figure 5.11G). By combining the western blots, T_1 analysis and 1D integrals, we can evaluate the overall sample integrity. The 1D data would indicate that the sample begins to break down almost immediately after sample preparation, however there is no evidence of this in the western blot analysis. Indeed, the origin of the continuously increasing signal at -119.2 ppm is not certain, and the overall integrity of the sample is therefore difficult to evaluate conclusively. The ^{15}N-SORDID experiments carried out at the beginning and the end of the ^{19}F experiments indicated that some degree of sample degradation had occurred (Figure 5.11F-H), although accurate analysis of the diffusion data, as mentioned before, proved challenging.

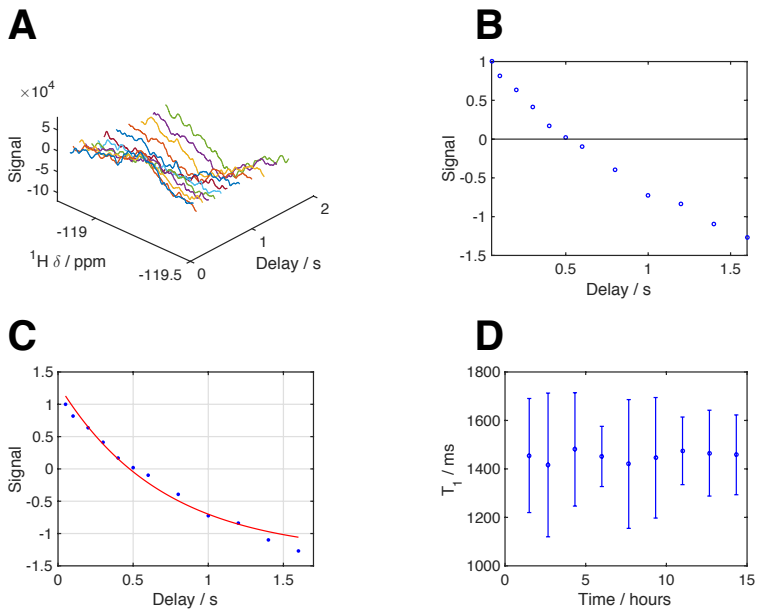

Figure 5.10: ^{19}F T_1 *inversion recovery measurements of the +31* ^{19}F *RNC.* **A:** T_1 inversion recovery data collected as soon as possible after placing the sample in the spectrometer. **B:** Intensity of the T_1 spectra shown in **A**. **C:** Fit of the intensities recorded in **B** to equation 1.7.9. **D:** ^{19}F T_1 values of the nascent chain over time. These were measured by repeating the analysis displayed in figures **A** to **C** on the nascent chain at incremental time points throughout the data acquisition period.

5.10 ^{1}H-^{15}N NMR analysis of the +67 domain 5 ^{19}F-^{15}N ribosome-bound nascent chain

The NMR experiments carried out on the +31 nascent chain were repeated for a 3-fluorotyrosine labelled sample of the +67 nascent chain. The sample was expressed and purified using an identical approach, as all FLN750 nascent chains can be purified in the same way.

The +67 nascent chain was prepared from a 2 litre culture, and after purification the sample was split into two equal halves. The first half was analysed by ^{1}H-^{15}N experiments on a 700 MHz spectrometer, while the second half was analysed entirely by ^{19}F experiments on the 600 MHz, fluorine capable spectrometer at Kent university.

In each case, the frozen sample was thawed on ice, and buffer exchanged into fresh Tico buffer (5 ×, 0.5 - 15 ml) to filter out nascent chain that can release in the freeze/thawing process. The final NMR sample (including 10 % D_2O and 0.01 % DSS) contained 7.4 µM ribosomes with a volume of 303 µl for the sample

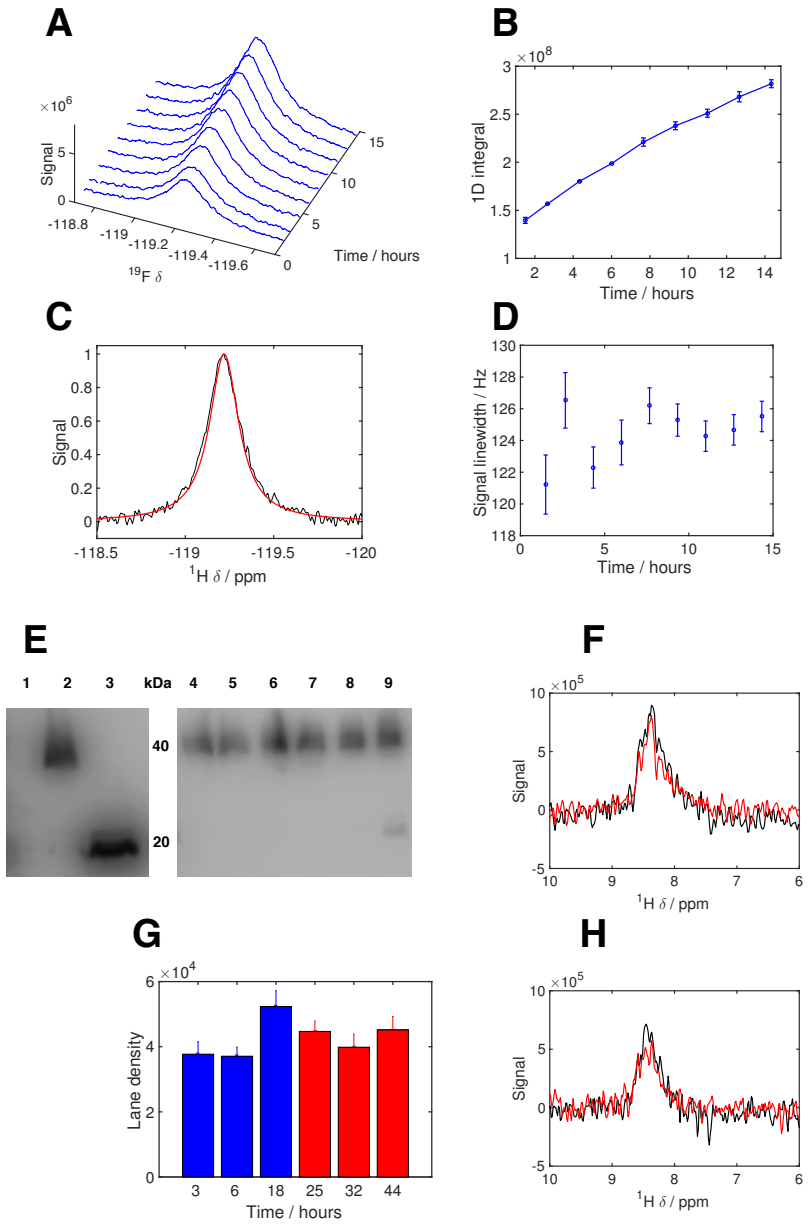

Figure 5.11: *Fluorine NMR spectra of the +31 RNC, and monitoring sample quality over time.* Monitoring of the ^{19}F NMR signal from the +31 nascent chain over time, as well as parallel biochemical experiments and SORDID diffusion to monitor sample breakdown over time. **A-B:** Appearance of the peak at -119.2 ppm in the ^{19}F spectrum, recorded over a total of 14 hours. The continual increase in the peak integral indicates that the sample is constantly changing with time. **C-D:** C shows the first 1D spectrum (from A) fitted to a lorentzian lineshape in red, showing that a lorentzian fits well to this peak. D shows how the linewidths measured from the ^{19}F peaks (shown in A) change over time **E and G:** Anti-his immunoblot analysis of the nascent chain, and quantification by densitometry. In G, blue bars correspond to samples taken when the sample was at 600 MHz, and red indicates when the sample was transferred to the 700 MHz spectrometer. **F and H:** Appearance of SORDID diffusion spectra at the beginning (F) and end (H) of acquisition of ^{19}F data on the 600 MHz spectrometer. Gradient strengths were 100 % (red) and 15 % (black) of G_{max}.

at 700 MHz, and with 9.8 μM with a volume of 304 μl for the sample at 600 MHz. Both samples were prepared in shigemi tubes.

5.10.1 ^{1}H-^{15}N NMR spectra of the +67 ^{19}F-^{15}N labelled nascent chain.

The ^{15}N spectra were recorded at 10 °C on a 700 MHz Bruker Avance spectrometer equipped with a cryogenic probe. ^{15}N filtered and ^{15}N edited spectra were initially recorded to estimate the extent of background ribosome labelling, and the nascent chain signal was estimated from a ^{15}N edited 1D ^{1}H spectrum. The sample was then analysed by recording ^{15}N SOFAST HMQC correlation spectra, ^{15}N edited SORDID diffusion spectra, and 1D ^{1}H spectra in an interleaved manner over several days, allowing its behaviour to be monitored systematically over time. The ^{1}H, SORDID and HMQC spectra took 3, 12 and 33 minutes to acquire respectively, giving a total of 48 minutes per cycle of experiments. This represents the approximate time resolution of the monitoring of sample degradation. The HMQC spectra are summarised in Figure 5.12, showing the +67 RNC that was collected in the present work (5.12A) alongside an HMQC of an unlabelled +21 RNC of the same FLN750 system (5.12B) (data provided by Anaïs Cassaignau). Investigation of the overlay of these spectra (5.12C) shows that the +67 is missing some key peaks corresponding to disordered domain 5, as expected for the folded nascent chain.

5.10.2 ^{19}F NMR spectra of the +67 ^{19}F-^{15}N domain 5 nascent chain

Although the nascent chain was deemed to be of good quality according to the amide signal and immunoblotting analysis, when its analogous sample was analysed by fluorine NMR, it was impossible to observe any fluorine signal, even after acquiring a single 1D experiment for 20 hours. As such the impact of fluorine on a structured nascent chain cannot yet be assessed at this stage, although it would appear that the reduction in tumbling that results from acquiring tertiary structure renders the fluorine linewidth too broad.

5.11 Ribosome binding experiments

If the ^{19}F spectra of the +31 RNC is overlaid with that of isolated domain 5 (FLN750 or FLN738), there is a huge chemical shift difference between the ribosome-bound and the isolated species (Figure 5.13) of over 15 ppm. This is a striking change in chemical shift to be observed. Hence, to gain a more clear understanding of the nature of the observed fluorine signal from the nascent chain, NMR spectra of isolated domain 5 samples were collected while in the presence of ribosomes.

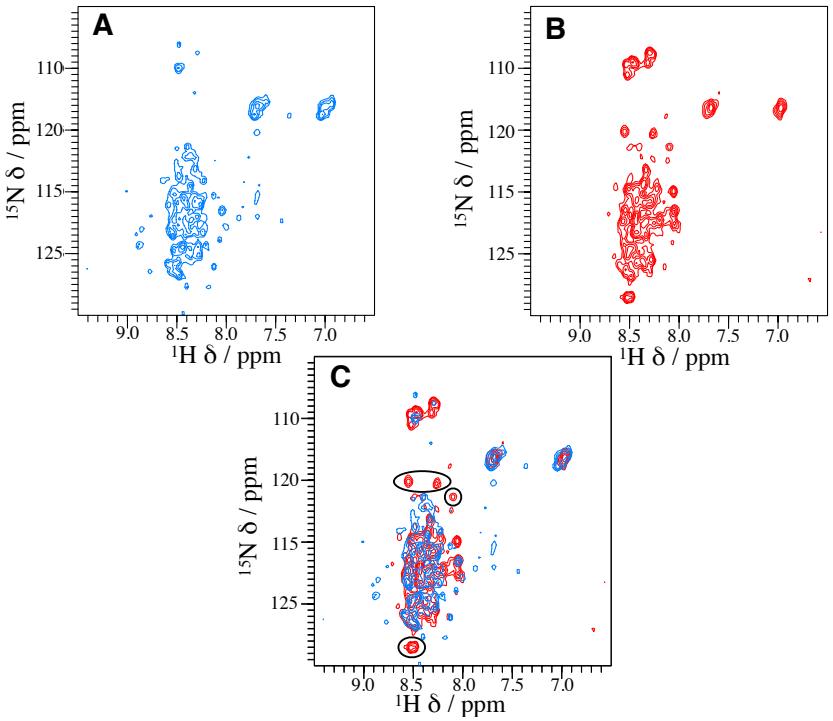

Figure 5.12: *^{1}H-^{15}N HMQC spectra of the +67 ^{19}F-^{15}N labelled +67 nascent chain.* **A:** HMQC of the ^{19}F-^{15}N labelled +67 domain 5 nascent chain. **B:** HMQC of the ^{15}N labelled +21 nascent chain (no fluorine). **C:** Overlay of both spectra, highlighting peaks missing from the +67 that correspond to unfolded FLN750. This indicates that the unfolded form of domain 5 is not present when domain 5 is stalled in translation 67 amino acids away from the PTC. Although we do expect the structured form of domain 5 to be present here, it cannot be observed using ^{15}N labelling as the signal is too broad.

NMR samples were prepared of FLN750 and FLN738 in isolation and in the presence of ribosomes. Samples were analysed by ^{1}H-^{15}N HSQC/HMQC and ^{19}F 1D spectra. For the ribosome titrations, each protein was mixed with ribosomes in a 1:1 ratio with 6 μM of protein against 6 μM of ribosomes.

Figure 5.13C-D shows ^{19}F spectra of FLN738 in isolation and in the presence of ribosomes. The ribosome introduces a small degree of line-broadening in the peaks centred at -137 ppm, suggesting some level of interaction. Of particular interest is the emergence of a peak at -119.2 ppm, indicating that an interaction between the unfolded domain and the ribosome leads to an enormous chemical shift change, and the new state of the protein appears to be similar to the state observed for the nascent chain. Alongside this, spectra were also acquired of FLN750 in the presence of ribosomes. However, this seemed to lead to a line-broadening so great that no fluorine signal could be observed. This suggests an interaction with the ribosome which is also enhanced by the presence of fluorine, since isolated, non-fluorinated FLN750 is known to have no interaction with the ribosome whatsoever [11].

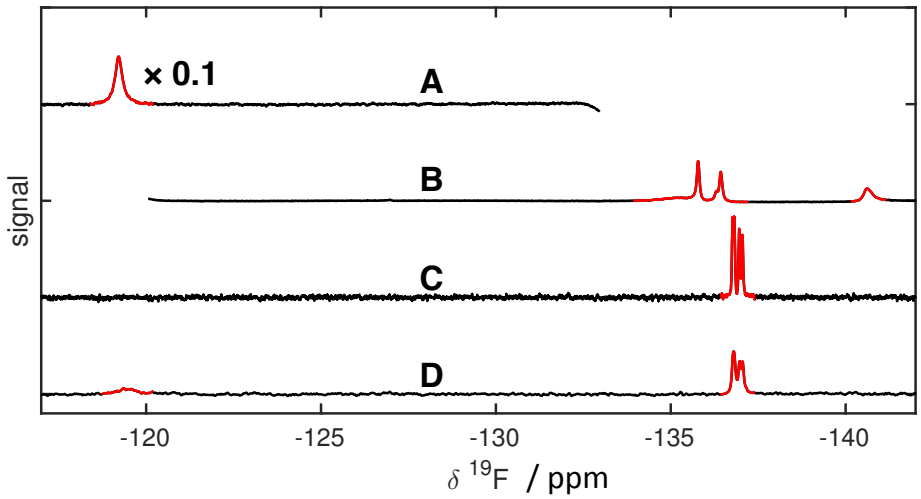

Figure 5.13: *^{19}F NMR spectra of the 3-fluorotyrosine labelled nascent chain, and of 3-fluorotyrosine labelled domain 5 in the absence and presence of ribosomes.* **A:** ^{19}F 1D NMR spectrum of the 3-fluorotyrosine labelled +31 RNC (scaled by 0.1) showing the presence of a strong peak at -119.2 ppm. **B:** ^{19}F1D NMR spectrum of isolated domain 5 (FLN750). **C:** ^{19}F1D NMR spectrum of isolated domain 5 that has been C-terminally truncated (FLN738). **D:** Spectrum of FLN738 in the presence of a 1:1 molar ratio of 70S ribosomes. All spectra were acquired at 10 °C and pH 7.5 on a Bruker Avance 600 spectrometer equipped with a cryogenic probe. The spectra as displayed have been normalised for concentration and number of scans.

Evaluation of peak integrals found the following surprising observation. After normalisation for number of scans and concentration, the integral of the peak at -119.2 ppm in the +31 RNC was found to be 106, relative to the integral of 43.4 for the isolated FLN750 peak, 25.2 for isolated FLN738, and 23 for FLN738 in the presence of ribosomes (including the peaks at -119.2 and -137 ppm). The origin of this discrepancy is uncertain, nevertheless it seems clear that the ribosome environment is responsible for a large change in the chemical environment, given its impact upon the chemical shift of fluorine.

We also investigated 3-fluorotyrosine in the presence of ribosomes. Figure 5.14 shows the results of this titration. What is immediately clear is that the large change in chemical shift that was observed in the FLN738 titration and on the nascent chain is also found here, suggesting that 3-fluorotyrosine may be influencing the interaction of the protein with the ribosome as opposed to revealing a native interaction. The spectra in Figure 5.14 also show a large baseline distortion. This distortion is due to the presence of fluorine in the hardware of the spectrometer, influencing the appearance of the spectra. Removing this distortion is sometimes possible by baseline corrections, as was the case for previous spectra, but in this case the distortion proved more difficult to eliminate.

To ensure that the signal originating at -119.2 ppm came from the interaction of protein or 3-fluorotyrosine

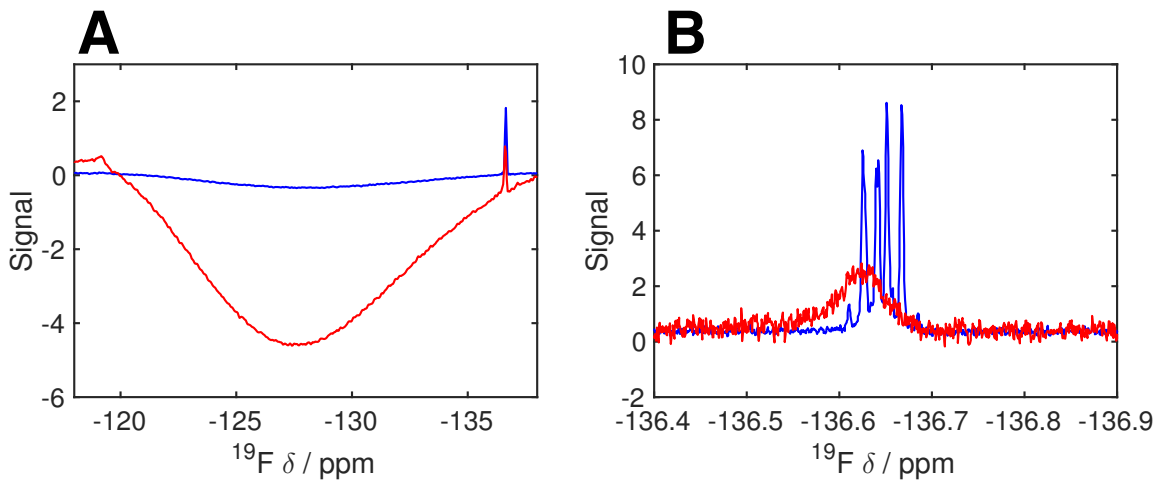

Figure 5.14: *^{19}F 1D spectra of 3-fluorotyrosine in the presence of ribosomes.* **A:** 1D ^{19}F NMR spectrum of 90 μM 3-fluorotyrosine in isolation (blue) and at 6.8 μM in the 1:1 presence of ribosomes (red). To reduce noise, these spectra were processed with 500 real points in the FID. This truncation reduces the detail in peak structure, but allows the small peak at -119.2 ppm to be observed clearly above the noise. **B:** The same spectrum as shown in A, zoomed in to show the peaks at -136.6 ppm. This spectrum was processed with no FID truncation, allowing the observation of greater detail in the fine structure of the peaks. Spectra were collected at 500 MHz on a bruker avance spectrometer equipped with a fluorine probe. The spectrum of isolated 3-fluorotyrosine was collected in 1024 scans, while the spectrum of 3-fluorotyrosine in the presence of ribosomes was collected with 19968 scans (ca. 18 hours). The broad line-width observed in the presence of ribosomes may be due to exchange broadening with the bound state of 3-fluorotyrosine.

with the ribosome, and not from background labelling of the ribosome itself, NMR spectra were carried out on ribosomes produced under conditions that maximised the chance of incorporation of the fluorinated analog (see Methods section 7.6.1). Even after acquisition of a ^{19}F 1D spectrum for over 18 hours, no signal was observed on these fluorinated ribosomes, regardless of the method used to produce them.

One of the most striking features of the peak at -119.2 ppm is that it can be observed at all. One might expect the high CSA of the fluorine nucleus to preclude from it being able to report on the ribosome-bound state of any species. Although this is not understood, it is extremely useful and promises to yield a great amount of new information about the role of the ribosome surface in co-translational folding.

Another striking feature is the large difference in chemical shift between the bound and free states of 3-fluorotyrosine. Again, the origin of this difference is not understood, but it may suggest that the ribosome surface acts as an extremely different environment from the rest of the surrounding solution, whether aqueous or cellular.

5.12 Discussion

The capacity of fluorine for investigating nascent chains is clear. The observation of an interaction between fluorine labelled domain 5 and the ribosome surface is of particular interest due to the observation of the ribosome-bound state of the protein. However, the significance of this state is unclear, since it appears to be the tyrosine itself that interacts with the ribosome, rather than a case of the tyrosine revealing an interaction with the nascent chain that was already present. However, the ability of the fluorine nucleus to be observed even when apparently bound to the 2.3 MDa 70S complex is an extremely important feature that may be invaluable for further studies, since to the best of our knowledge there is no precedent for the direct observation by NMR of the ribosome-bound state of any molecular species. The large change in chemical shift is also a unique result, which indicates that the ribosome surface is a highly unique environment that has extreme effects upon the chemical environment of the species it interacts with. The extreme nature of the environment is also hinted at by the interaction between α-synuclein and the ribosome, where very strong line-broadening of amide proton signal is observed. Further to this, the ability of the ribosome to cause precipitation of numerous proteins is again suggestive of a highly unique environment that can adversely affect the stability of a protein outside the context of the cellular environment. The ability of fluorine to therefore investigate this environment where more conventional nuclei cannot makes it potentially a very powerful nucleus, and we hope that its further use in the investigation of ribosome bound nascent chains will provide unprecedented insight into the mechanism of co-translational folding and nascent chain/ribosome interactions.

Chapter 6

Conclusions

In conclusion, with the lineshape simulations shown in Chapter 2 it was possible to provide theoretical insight into the limitations and advantages of the slow and fast exchange regimes in the context of a ribosome-bound nascent chain. Specifically, we saw that in an appropriate context, fast exchange chemical shift timescales may be able to provide more detailed information about the folding behavior of an RNC than has been observed for the slow exchanging systems that have currently been developed, such as the domain 5 protein. These simulations could be expanded upon, with a deeper investigation of different exchange rates and transverse relaxation rates. Another approach would be to determine if it is possible to detect changing kinetics. For example, as exchange kinetics near the mid-point of folding of a protein are typically slower, one could replicate this effect in a simulation to observe how this affects the spectra. The difficulties in this approach are that one very quickly starts producing large quantities of data, which takes up time and disk space, and the effective analysis of such data can be very challenging. However, with the appropriate simulation parameters, this method can help to give a strong proof of principle for a potential approach, before it is embarked upon experimentally.

From the simulations, we determined that fast folding proteins may have an advantage that could be exploited on the ribosome by NMR. Two candidate nascent chains - the HP36 and the GA module - were therefore selected as they are known to have extremely fast folding kinetics in isolation. They are now currently in the process of development towards testing the hypothesis posited regarding the advantages of fast exchange for studies of folding on the ribosome.

Initial kinetic studies on these proteins (in Chapter 3) proved problematic, as it appeared that at room temperature we could not reproduce the fast exchange kinetics that had been observed in the literature. This

is probably due to the combination of the effects of urea and the lower temperatures used in this thesis than those used in the literature. Nevertheless, given that it is very difficult to predict the effect of the ribosome itself on folding kinetics, it was decided that the HP36 and the GA module are strong candidates for the exploration of fast exchange in this context.

Although the HP36 proved extremely difficult to work with as a nascent chain, some success was obtained with the GA module, as it was possible to express and purify a GA module ribosome-bound nascent chain in quantities sufficient for NMR spectroscopy (Chapter 4). We were not able to subsequently observe the GA module on the ribosome, however it is very likely that with further development this protein will serve as a powerful complement to the slow folding domain 5 nascent chain. We also expect that further work on successful expression of the HP36 should yield similar progress as was obtained for the GA module. These two proteins promise to lead to exciting new insights into the nature of co-translational folding as studied by NMR spectroscopy.

The impact of the ribosome surface was striking when considering its interactions with numerous proteins, including the HP36 and the GA module. Although such interactions are problematic, as they lead to line-broadening effects, their potential relevance to the phenomenon of co-translational folding cannot be ignored. Observations of the impact of the histidine purification tag indicate that electrostatics might play a role in the interaction of proteins with the ribosome surface. However, there is also evidence of a hydrophobic driven effect. The addition of arginine to NMR buffers lead to a great reduction in the extent of line-broadening observed for the GA module and the HP36 in the presence of ribosomes. Given that arginine has been shown to aid in solubilisation of hydrophobic regions of proteins, this indicates a potentially hydrophobic driven mechanism of interaction between ribosomes and proteins. Mutational studies on the GA module and the HP36 would be very useful here, to probe the impact of electrostatics or hydrophobicity though appropriate mutations. Repetition of the ribosome binding experiments described in Section 3.5 with such mutations might help to provide some insight into the impact of hydrophobicity/electrostatics on the nature of the protein:ribosome interaction. This would complement the work by Knight et al. [3] and Deckert et al. [53], both of which focussed on understanding the nature of the interaction of nascent chain with the ribosome surface.

Some understanding of the interaction between protein and ribosome was gained through the studies of α-synuclein in Chapter 5. From these titrations, combined with relaxation measurements, it was shown that a simple 1:1 interaction model fits well to the data we currently have on the interaction between α-syn and the ribosome particle. This model indicates that the interaction between α-syn and the ribosome is weak

and transient, but since it occurs extremely rapidly, this leads to an accumulation of line-broadening effects that render the bound-state unobservable. The extremely rapid kinetics might suggest a highly non-specific interaction, as opposed to binding to a specific area of the ribosome, and since the α-syn was free, it had access to the entire ribosome surface in a way that a nascent chain would not. However, this may still have some biological relevance, and further work on the interaction, combined with studies of the nascent chain, will help to elucidate this further.

Perhaps one of the most surprising results was the direct observation of what appeared to be a ribosome-bound species through the ^{19}F nucleus (at the end of Chapter 5). For ^{1}H, ^{15}N and ^{13}C, the ribosome-bound chemical shift of these nuclei cannot be observed. However, as shown in Figure 5.13 and Figure 5.14, whenever 3-fluorotyrosine is in the presence of 70S ribosomes, a very broad low field shifted peak emerges. This was observed on the ^{19}F labelled nascent chain, as well as for isolated 3-fluorotyrosine.

The capacity of the fluorine nucleus to to provide direct observation of large, fast relaxing species is likely to be an extremely powerful addition to the toolkit of NMR spectroscopy for the study nascent chains. In particular, ^{19}F NMR may be extremely valuable for in-cell NMR studies of RNCs [142], where uncluttered signal, and observation of very fast relaxing species, is extremely important for reliable studies. However, there were some problems regarding the efficiency of fluorine labelling. In particular, further work to characterise the ^{19}F labelled sample of domain 5 that was produced will need to be carried out, since it is unclear how effective the labelling process was for this protein. It is also very important to subsequently optimise the method used to produce fluorine labelled samples, to ensure that high quality, fluorine labelled samples can be reliably produced.

Nevertheless, there is some evidence that fluorine signal could be observed on RNCs, and the effect of the ribosome on the chemical shift of fluorine in 3-fluorotyrosine indicates that the ribosome surface acts as an environment markedly different to the surrounding aqueous solution. This is supported by the line-broadening effects on other nuclei, and the tendency of ribosome solutions to induce precipitation of some proteins. We can therefore conclude that the ribosome is clearly having an effect on proteins as they are translated, and the fluorine nucleus may become an ideal choice for detailed investigation of such interactions if it is able to show the directly ribosome-bound state of a molecular species.

6.1 Future work on co-translational folding

Despite its remarkable power and versatility, NMR spectroscopy will always be limited by the correlation time of the ribosome, making the observation of species near to the centre of the ribosome almost impossible. The combined results of cryo-electron microscopy [45, 143] and x-ray crystallography [57] have been instrumental in determining the structure of the core of the ribosome, but it was only by combining these results with NMR studies that it was possible to also obtain structural information on the flexible outer-regions of the 70S particle [50].

In the future, it will be the combination of data gathered from all three of these extremely powerful techniques, on high quality samples of translation stalled products, in conjunction with computational studies, that will give the greatest level of insight into the structure and folding behaviour of proteins as they are bound to the ribosome, from the PTC to the surface.

As well as this, there are investigations to be made into real time studies, using more rapid techniques such as dynamic fluorescence depolarisation [144]. Although the technology at the time of writing has not yet achieved this, it may be feasible that a technique such as fluorescence depolarisation could be used on a protein translating on the ribosome in real time. One could even circumvent the problem of inhomogeneous translation by performing single molecule fluorescence studies, [145] thus providing unparalleled time resolution on the translation process.

Almost all studies of co-translational folding, both here and in the bulk of the literature, have been carried out on *E. coli* ribosomes. These are the easiest to purify and the most well understood. However, co-translational folding as it occurs on eukaryotic ribosomes, and in particular, the human ribosome, can be viewed as a logical end-point of many of the efforts described here and elsewhere. Again, the technology for the purification of eukaryotic ribosomes does not produce samples in sufficient yield for much more than crystallographic studies [146], and although this has provided a great wealth of information on the complexity and structure of such species, it is not at the level of detail that has been obtained for prokaryotes.

In recent years, eukaryotic nascent chains have in fact begun to be investigated. Zhang et al. successfully investigated model proteins as RNCs in a yeast translation system, using UV crosslinking and biochemical assays [147]. This signifies the emergence of a new revolution in the study of co-translational folding. Hence, all the work that has been carried out on the systems described in this thesis, and in the literature, on prokaryotic ribosomes, could well be repeated on their eukaryotic counterparts. This represents a general progression towards a more realistic description of co-translational folding in a human-physiological context, using human ribosomes and human proteins, to truly understand how the proteins in our bodies adopt the

structures required for their function.

Chapter 7

Materials and Methods

7.0.1 NMR spectra

Spectra were acquired on a range of different spectrometer frequencies. Due to the wide range of criteria behind each individual experiment, these details are provided in the results section alongside each relevant experiment. Spectrometers used included the frequency range of 500, 600, 700 and 800 MHz, with the use of cryogenically cooled probes. The 500 and 600 MHz spectrometers were also equipped with a fluorine probe. All spectrometers were bruker instruments, and data processing was carried out in NMRpipe. Further analysis was also carried out in matlab, where stated [67]. All spectrometers were Bruker Avance instruments.

Sample preparation

All work involving living samples was carried out in the presence of a yellow flame, using preparatory reagents sterilised by autoclaving or filtering through a 0.22 μm filter. Unless otherwise stated, all reagents were obtained from Sigma-Aldrich. All centrifuging equipment was purchased from Beckmann Coulter Genomics.

7.1 Production and purification of protein samples

7.1.1 Plasmid transformation

1 μl of recombinant plasmid DNA containing the gene of interest was added to a 25 μl aliquot of competent BL21 DE3 (gold) *E. coli* cells. Transformation was induced by incubation on ice for 30 mins, followed by a 45 second heat shock at 42 °C. Addition of 950 μl of LB medium was followed by aeration at 37 °C (30-60

mins). The resulting culture was centrifuged (tabletop centrifuge, 13,000 rpm, 4 mins) and the supernatant discarded. The pellet was resuspended in the LB that remained of the supernatant, spread onto an LB agar plate (1% ampicillin or kanamycin, as required) and incubated at 37 °C (16 hours). Each colony from the overnight plates could then be used to inoculate further growth media for expression of the protein of interest.

7.1.2 Standard growth and expression media

All protein samples were produced using standard *E. coli* expression techniques. For expression of unlabelled proteins, transformed cells were grown in LB medium (5 g l^{-1} yeast, 10 g l^{-1} NaCl, 10 g l^{-1} Tryptone) with appropriate antibiotics.

For isotopic labelling, an LB preculture was used to inoculate a ^{15}N and/or ^{13}C supplemented M9 medium for expression. M9 salts are added as a 10 × stock (See below for recipes).

Other stock solutions included an EM9 medium for expression of nascent chains, a non-inducing MDG medium for growing cultures to high cell densities, and a trace metals solution;

M9 medium:

1× M9 salts, 4 g l^{-1} glucose (2 g l^{-1} if ^{13}C labelled), 1 g l^{-1} $^{15}NH_4Cl$, 100 μM $CaCl_2$, 2 mM $MgSO_4$, 10 ml l^{-1} BME vitamins (Sigma), 100 mg l^{-1} ampicillin (or 30 mg l^{-1} kanamycin).

10× M9 salts:

30 g l^{-1} KH_2PO_4, 67.9 g l^{-1} Na_2HPO_4, 5 g l^{-1} NaCl, pH 7.5

EM9 medium:

1× EM9 salts, 1 g l^{-1} $^{15}NH_4Cl$, 4 g l^{-1} glucose, 2.5 ml l^{-1} BME vitamins (Sigma), 200 μM $CaCl_2$, 5 mM $MgSO_4$, 0.25× trace metals

10× EM9 salts:

0.5 M Na_2HPO_4, 0.25 M KH_2PO_4, 10 M NaCl, pH 8.0-8.2

MDG medium:

1× MDG salts, 0.25 % (w/v) L-aspartic acid, 2 mM $MgSO_4$, 0.2 × trace metals, 4 g l^{-1} glucose, 100 mg l^{-1} ampicillin (or 30 mg l^{-1} kanamycin).

25× MDG salts:

0.625 M Na_2HPO_4, 0.625 M KH_2PO_4 1.25 M NH_4Cl, 0.125 M Na_2SO_3

Trace metals (1000× stock):

50 mM $FeCl_2$ (dissolved in 0.1 M HCl), 20 mM $CaCl_2$, 1 mM $MnCl_2{\cdot}4H_2O$, 1 mM $ZnSO_4{\cdot}7H_2O$, 2 mM

each ($CoCl_2 \cdot 6H_2O$, $CuCl_2 \cdot 2H_2O$, $NiCl_2 \cdot 6H_2O$, $Na_2MoO_4 \cdot 2H_2O$, $Na_2SeO_3 \cdot 5H_2O$, H_3BO_3), all filtered (0.22 μm).

A 1000× stock solution of rifampicin was also prepared, at 150 mg ml^{-1} in DMSO, for expression of ^{15}N labelled RNC samples against an unlabelled background.

7.1.3 Expression of ^{15}N labelled α-synuclein

A single colony of transformed BL21 cells containing the α-syn or H_6-αsyn gene was added to LB (5ml, 1% ampicillin) and aerated for 5 hours (250 rpm, 37°C). 100 μl of this pre-culture was added to 100 ml LB medium (1% ampicillin) and aerated (37 °C, 250 rpm, 16 hours, $OD_{600} \approx$ 2-3). The pre-culture was centrifuged at high speed for 10 minutes, and the resulting pellet resuspended in M9 medium (50 ml, no nitrogen source). To clean the pellet of remaining LB, the sample was centrifuged and resuspended twice more before resuspension in 2 litres of M9 medium (0.1% (w/v) $^{15}NH_4Cl$), to an OD_{600} of 0.1. Aeration continued at 37 °C and 250 rpm until the OD_{600} reached 0.6. Expression was induced by addition of 1 mM IPTG followed by further aeration for 4 hours. The final culture was harvested by centrifugation at 4,000 rpm for 20 minutes in a JLS 8.1 rotor, and stored at -20 °C.

7.1.4 Purification of α-syn and H_6-αsyn

Cells containing expressed α-syn were thawed and resuspended in a lysis buffer (100 mM Tris-HCl, 10 mM EDTA, 2 mM BME, protease inhibitors (Roche, 1 tablet per 50 ml), pH 8.0). Cells were lysed by two liquid nitrogen freeze thaw cycles and subsequent sonication (6 cycles, 30 seconds on/30 seconds off). Cellular debris was pelleted by centrifugation (18,000 rpm, 30 minutes, SS34 rotor) and the remaining supernatant was boiled for 20 minutes. The lysate was cleaned again by centrifugation (30 minutes, 13,500 rpm), and the supernatant stirred with streptomycin sulphate (10 mg for 1 ml of supernatant, 4 °C, 20 minutes). Another centrifugation step was followed by precipitation of the protein with ammonium sulphate (400 mg ml^{-1}, 4 °C, 30 minutes). After a final centrifugation step, the supernatant was discarded and the pellet resuspended in $<$ 5 ml of buffer A (25 mM Na_2HPO_4, 150 mM NaCl, 2 mM β-mercaptoethanol, pH 7.5), before dialysis against buffer A (4 litres, 4 hours). This dialysis step was repeated to ensure complete exchange into buffer A.

The sample was then purified by size exclusion on an S75 16/600 gel filtration column equilibrated in buffer A. The eluent was collected in 2 ml fractions, and the fractions containing the pure protein, as evaluated by SDS-PAGE, were pooled and concentrated in a 5 kDa centrifugal cutoff filter. Purity was

assessed again by SDS-PAGE, and concentration was determined from the absorbance at 260 nm (A_{260}).

The purification of H_6-αsyn was identical to that of the wild type protein, however an additional nickel column step was added prior to size exclusion;
After dialysis, the sample was bound to an NTA nickel resin purchased rom Qiagen (2 hours, 4 °C). The resin was washed with buffer A (250 ml), and Bradford reagent was used to assess when all unwanted proteins had been cleaned from the resin. The protein was then eluted by addition of 150 mM imidazole to the buffer, and eluent was collected until the Bradford reagent indicated elution had stopped (150 ml). Final yield was 0.5 mM in 1 ml from 1 litre of culture of H_6-αsyn, as determined by the OD_{280}.

7.1.5 Purification of GA module

The GA module was expressed by addition of 1 mM IPTG to bacterial growth at an OD_{600} of 0.6 and incubation at 37 °C for 6 hours. Cells were harvested by centrifugation and stored at -20 °C.

Frozen cells were resuspended in lysis buffer (tris, pH 7.5, 20-30 ml per litre of growth). The cells were lysed by 3 liquid nitrogen freeze thaw cycles, followed by sonication (8 cycles, 30 seconds on, 30 seconds off). After centrifugation at 18,000 rpm (45 minutes, SS34) the supernatant was loaded onto a nickel resin equilibrated in tris buffer and left to bind for 2 hours. The resin was washed with 200 ml of tris buffer, until bradford assay indicated that all impurities had been washed away, and the sample was eluted with 250 mM imidazole in tris buffer (pH 8.0, 100-150 ml). Immediately following elution, the sample was concentrated to less than 5 ml and loaded onto a 16/600 S75 gel filtration column (GE healthcare). The GA module eluted at 94 ml. Elution was collected in 1.5 ml fractions, and purity assessed by SDS-PAGE before pooling and concentration of the most pure fractions. At this stage, the sample could be stored at -80 °C.

TEV cleavage of the GA module

To remove the his-tag, the GA module was incubated in a molar excess of TEV protease for 16 hours at 4 °C. The resulting solution was then bound to a nickel column for two hours, and the flow through collected. The flow through contained only the cleaved GA module. This was further purified by another size exclusion step to remove residual impurities.

7.1.6 Expression and purification of NTL9-HP36 fusion protein

The pET3a-NTL9-FXa-HP36 plasmid was transformed into competent BL21 DE3 *E. coli*. A preculture in LB medium was used to inoculate a ^{15}N labelled M9 medium which was grown at 37 °C and 250 rpm until

reaching an optical density of 1.5. Expression was then induced by addition of 1 mM IPTG and incubation for four hours at 37 °C. Cells were harvested (4000 rpm, 20 minutes) and either stored at -20 °C or immediately lysed by 2 freeze thaw cycles and sonication (8 × 30 seconds on/off) after resuspension in lysis buffer (20 mM Tris pH 7.5). The insoluble fraction was removed by centrifugation (18,000 rpm, 45 minutes, SS34) and loaded onto an ion exchange SP sepharose column. The fusion was eluted in a 0-2M NaCl gradient and further purified by gel filtration on a 16/600 S75 column.

Cleavage of fusion protein using factor Xa

The fusion protein was incubated in the presence of factor Xa for 16 hours at 24 °C in a cleavage buffer (50 mM Tris-HCl, 100 mM NaCl, 5 mM $CaCl_2$, pH 8.0). 2 units of factor Xa were added for each milligram of protein. After cleavage, the protein was purified by HPLC on a C8 preparative column. The cleaved protein was eluted in a linear gradient from buffer A (0.1 % TFA in H_2O) to buffer B (90 % acetonitrile, 99.9 % H_2O, 0.1 % TFA). The gradient was applied over 100 minutes after initially washing the column in 100 % buffer A for 10 minutes. Both the NTL9 and HP36 containing fractions were pooled separately and dried on a rotary evaporator. The solid was then stored at -20 °C.

7.1.7 Expression of ^{19}F labelled α-synuclein

Expression via glyphosate approach

An LB preculture of BL21 cells transformed with the α-syn plasmid was used to inoculate an M9 medium composed of the required isotopic components (^{15}N). This was incubated (37 °C, 200-250 rpm) until it reached an OD_{600} of 0.4. At this stage, the medium was supplemented with glyphosate (500 mg l^{-1}), D-L 3-fluorotyrosine (35 mg l^{-1}), L-phenylalanine (30 mg l^{-1}) and L-tryptophan (30 mg l^{-1}). This was allowed to incubate for a further 30 minutes (OD_{600} = 0.6-0.7), before induction by addition of 1 mM IPTG. The flask was then incubated at 37 °C for 16 hours before harvesting of cells. The sample was purified using the method described in section 7.1.4.

7.1.8 Expression and purification of ^{19}F labelled FLN750

Expression via glyphosate approach

An LB preculture of BL21 cells transformed with the domain 5 plasmid was used to inoculate an M9 medium composed of the required isotopic components (^{15}N). This was incubated (37 °C, 200-250 rpm)

until it reached an OD_{600} of 0.4. At this stage, the medium was supplemented with glyphosate (500 mg l^{-1}), D-L 3-fluorotyrosine (35 mg l^{-1}), L-phenylalanine (30 mg l^{-1}) and L-tryptophan (30 mg l^{-1}). This was allowed to incubate for a further 30 minutes (OD_{600} = 0.6-0.7), before induction by addition of 1 mM IPTG. The flask was then incubated at 37 °C for 16 hours before harvesting of cells.

Expression via auxotroph approach

A bacterial strain auxotrophic for tyrosine known as C43(DE3) ML14 [148–150] was transformed with the FLN750 plasmid. A preculture in LB medium was prepared, and used to inoculate an EM9 medium supplemented with L-tyrosine (3.6 mg l^{-1}) and D/L-3-fluorotyrosine (36 mg l^{-1}). Expression was induced with 1 mM IPTG when the cells reached an optical density of 0.65 and incubated for 16 hours at 37 °C. This process of expression was repeated in an identical manner for the FLN750 and FLN738 constructs. The ML14 auxotrophic strain was kindly donated by Dr Toshio Iwasaki of Nippon Medical School.

7.1.9 Purification of the FLN750 and FLN738 proteins.

Cells were resuspended in lysis buffer (20 mM Tris pH 7.5) and lysed by 2 liquid nitrogen freeze thaw cycles and sonication (8 cycles, 30 seconds on/off). The soluble fraction was isolated by centrifugation (18,000 rpm 45 minutes, SS34), and the supernatant loaded onto a nickel IDA resin equilibrated in Tris pH 7.5. Elution was induced with 250 mM imidazole, and the subsequent elution was purified further by size exclusion (16/600 S75, GE healthcare, 20 mM Tris pH 7.5). The purification process was identical for the FLN750 and FNL738 proteins.

7.1.10 Expression and purification of TEV protease

A triple mutant variant of the H_6-TEV-MBP construct was grown in an autoinduction medium (TB (60 g l^{-1}), glycerol (10 ml l^{-1}), 1 % ampicillin) for 8 hours at 37 °C, until growth saturation was obtained. The temperature was then reduced to 30 °C and incubation continued for 14 hours before harvesting the cells.

Purification of the protease

Cells were suspended in a lysis buffer (25 mM Na_2PO_4, 20 % (v/v) glycerol, 500 mM NaCl, 20 mM imidazole, 1 mM BME, 100 mg lysozyme, 1 tablet Roche protease inhibitors, pH 7.5, 5 ml per gram of cells) and incubated at 4 °C with rotation for 20 minutes. The cells were lysed by sonication (6 cycles, 30 seconds on/off) whilst keeping them chilled on ice and the cellular debris was pelleted by centrifugation (18,000

rpm, 45 minutes, SS34 rotor). The soluble fraction was filtered through a 0.45 μm filter before being loaded onto a HisTrap HP column (GE Healthcare) equilibrated in buffer A (as lysis buffer, minus lysozyme and inhibitors).

The column was washed with a combined flow of 98 % buffer A and 2 % buffer B (as buffer A, plus 500 mM imidazole). The TEV was eluted in a gradient from 2 % to 100 % buffer B over 20 column volumes (100 ml), collecting 2 ml fractions. Fractions were checked for the presence of TEV (30 kDa) and MBP (45 kDa) by SDS-PAGE. Fractions that contained more TEV than MBP were pooled and twice dialysed into a dialysis buffer (25 mM Na_2PO_4, 200 mM NaCl, 10 % (v/v) glycerol, 5mM BME, pH 7.5) (1 litre, 4 °C for 16 hours or 25 °C for 4 hours).

The dialysed TEV was concentrated down to 5 ml, or until aggregation started to occur, at which point excess TEV was frozen in liquid nitrogen and stored at -80 °C. The remaining TEV was then filtered (0.45 μm) and purified on a 16/600 S75 gel filtration column (GE healthcare) equilibrated in a size exclusion (SEC) buffer (as dialysis buffer, filtered (0.22 μm) and degassed). The elution was collected in 1.5 ml fractions. Fractions were evaluated for purity by SDS-PAGE, and pure fractions were pooled and concentrated to between 5 and 20 mg ml^{-1} of TEV. The protein was split into 50 μl aliquots, snap frozen and stored at -80 °C. The concentration of TEV was evaluated by measuring the absorbance at 280 nm ($A_{280} = 1.126 \implies 1$ mg ml^{-1}).

7.2 DNA manipulation techniques

7.2.1 Site directed mutagenesis and PCR amplification

All mutagenesis and amplification reactions were carried out using a KOD hot start DNA polymerase kit purchased from Novagen.

Mutagenesis reactions were prepared by mixing 1 μl of template DNA with $MgSO_4$ (25 mM, 1.5 μl), dinucleoside triphosphates (dNTPs, 2 mM, 2.5 μl), forward and reverse primers (10 μM, 0.4 μl) KOD buffer (10 ×, 2.5 μl) and KOD DNA polymerase (0.5 μl). autoclaved water was added to the mixture to make a final volume of 25 μl.

The PCR cycle began with 2 minutes of denaturation at 95 °C, followed by 20 repetitions of a denature, anneal, elongation cycle (95 °C 2 minutes, 55 °C 10 seconds, 70°C 3 minutes) and a final elongation step (70 °C, 5 minutes) before reducing temperature to 25 °C. 5 μl of PCR product was used to confirm reaction success by electrophoresis on an agarose gel, and parent DNA was digested by addition of the enzyme dpnI

(New England Biolabs, 1μl), 10 × cutsmart buffer (2.5 μl) and water (1.5 μl) to the remaining 20 μl of product. This was incubated at 37 °C for 1 hour, and 1 μl of the product was transformed into DH5α cells, which were grown on LB plates in presence of the appropriate antibiotics. Single colonies were grown in LB, miniprepped (DNA purification kits from Qiagen) and DNA was sequenced by New England Biolabs (NEB).

Colony PCR

For rapid identification of products from ligation reactions (see Section 7.4.3), colony PCR using Taq DNA polymerase (NEB) as the elongation enzyme was used to directly determine if colonies of transformed cells contained the intended plasmid. A reaction mixture was prepared containing 10 × Taq buffer (2.5 μl), forwards and reverse primers (10 μM, 1.25 μl), dNTPs (2 mM, 2.5 μl), and Taq DNA polymerase (0.5 μl) in a total volume of 25 μl of autoclaved water. A clean tip that had been scraped over a cell colony was briefly suspended in the reaction mixture to inoculate it with bacteria. The rection cycle consisted of an initial 95 °C denaturation step, followed by 30 repetitions of a denaturation, annealing, elongation cycle (95 °C1 minute, 55 °C 1 minute, 72 °C 0.5 minutes) and a final elongation step (72 °C2 minutes). Reaction success was checked by gel electrophoresis.

7.3 Production of nascent chains

7.4 Cloning

To produce plasmids capable of expressing ribosome-stalled nascent chains, a combination of site directed mutagenesis and DNA digestion/ligation techniques were used to modify existing plasmids to incorporate the proteins of interest to this study.

7.4.1 GA module cloning

A kanamycin resistant plasmid capable of expressing the GA module was kindly donated by Dr Sarah Lejon. The N-terminus contained a thrombin cleavable hexahistidine tag followed by an NdeI digestion site (codon sequence CATATG; amino acid sequence HM);
MGSSHHHHHHSSGLVPRGSHMTIDQWLLKNAKEDAIAELKKAGITSDFYFNAINKAKTVEEVNALK
NEILKAHA

The thrombin site (LVPRGS) was replaced with a a TEV cleavage site by mutagenesis. A SacI site (GAG

CTC; EL) was inserted at the C-terminus. The NdeI site was then removed and replaced beyond the his-tag, so that the final sequence at this stage was as follows;
MHMSHHHHHHSSGENLYFQGTIDQWLLKNAKEDAIAELKKAGITSDFYFNAINKAKTVEEVNALKN EILKAHAEL-stop.

This sequence translates to the following; NdeI-H6-TEV-GA-SacI. The final product was amplified by miniprep from 5 ml of DH5α in LB (16 hours, 37 °C).

7.4.2 Digestion of modified GA module and nascent chain plasmids

40 ul of DNA plasmid with an estimated concentration of 400 ng μl^{-1} was mixed with the SacI restriction digestion enzyme (3 µl, 10 units μl^{-1}) in the presence of 1 × BSA and 1 × cutsmart buffers, with a total volume of 100 µl. The mixture was incubated at 37 °C for 3 hours.

After digestion, the plasmid was purified on a resin and eluted in 40 ul of tris pH 8.0. The digestion step was then repeated on the clean DNA sample, replacing SacI with NdeI, cleaving the GA module fragment from the plasmid DNA. This fragment was isolated by electrophoresis on a 1 % agarose gel. the GA module fragment was then excised from the gel and purified using a gel excision kit.

To generate the RNC plasmid, it was necessary to then produce a complementary vector into which the GA module insert could be placed. For this, the dom5+110 RNC plasmid was used;

NdeI-H6-dom5-SacI-dom6(+110)-SecM

This was also treated to the same NdeI SacI digestion ligation procedure. At the excision stage, the larger fragment, known as the vector was isolated and purified by electrophoresis and excision.

7.4.3 Recombination of GA module fragment and nascent chain vector by ligation

The concentrations of the GA module Insert and plasmid vector were estimated by the absorbance at 260 nm. The size of the insert was calculated to be 0.19 kb, while the vector was estimated to be 9 kb, by inspection of the electrophoresis gel. From this, the molar concentration of vector and insert was estimated.

For the ligation, four molar concentration ratios of insert:vector ratio were trialled; 3;1, 5:1, 10:1 and 20:1. Each ligation reaction was composed of 100 ng of insert in the presence of the appropriate quantity of vector. Final reaction volume was 20 µl in 1 × DNA ligase buffer and 1 µl T4 DNA ligase (400 units).

Alongside these reactions, 3 control reactions were also prepared; vector control 1 (no insert), vector control 2 (no insert, no ligase), and insert control (no vector). All reactions were incubated at 16 °C for 16

hours (or 25 °C for 1 hour) followed by heat deactivation at 65 °C for 20 minutes. The reactions were then chilled on ice.

5 µl of each ligation reaction was transformed into 25 µl of DH5α *E. coli* cells and grown on LB-Ampicillin plates. The insert control was grown on an LB-kanamycin plate since the insert came from a kanamycin resistant plasmid. The success of the reaction was confirmed when the control reactions gave relatively few or no colonies on their respective LB plates, while the true reactions did yield colonies. When this result was obtained, 6 colonies were selected at random from the successful reaction plates and used to inoculate 5 ml of LB media (ampicillin) and to initiate a colony PCR reaction (see Section 7.2.1), using as the primer a sequence that recognises the GA module.

The products of the PCR reaction were investigated by electrophoresis, and the reactions that strongly yielded product were identified as likely to contain the desired plasmid. These reactions were hence tested by DNA sequencing of the miniprepped product of the respective LB growth. Successful reactions were stored at -20 °C.

7.4.4 Production of RNC plasmid for the HP36 nascent chain construct

The HP36 RNC construct was produced using a similar method as described for the GA module RNC as described above, however the vector used was the α-1 antitrypsin (AAT) +110 RNC instead of the dom5+110 RNC. This construct contains a TEV cleavage site already between the N-terminal his-tag and the start of the AAT sequence, allowing its incorporation onto the HP36 RNC. The NTL9-HP36 fusion construct and the AAT RNC construct were as follows;

NTL9-Xa(cleavage)-HP36

H6-TEV-AAT-SacI-dom6+110-SecM-Stop

In order to remove the AAT sequence from the vector, an NheI site was amplified in between the TEV and and AAT encoding sequences, allowing the vector to be used to generate the constructs shown in section 4.1.

7.5 Large scale growth and expression of ^{15}N labelled ribosome-bound nascent chains

A single colony of competent BL21 DE3 transformed with the required RNC plasmid was used to inoculate a 5 ml culture of LB medium (+1 % ampicillin). This was incubated at 37 °C for 4-8 hours until it was almost opaque with growth. 500 µl of this culture was used to inoculate 500 ml of a non-inducing MDG medium

(1 × MDG salts, 0.25 % (w/v) L-aspartic acid (from 5 % w/v stock, pH 7.0, autoclaved), 2 mM $MgSO_4$, 0.2 × trace metals, 0.4 % w/v glucose, 100 mg l^{-1} ampicillin). This medium was incubated for 16-18 hours (overnight) at 37 °C and 250 rpm, until the stationary phase of growth was reached (OD_{600} 3-4). Once saturation was reached, the cells were centrifuged (3000 rpm, 25 minutes, 4 °C Beckmann JLA 8.1 rotor). The cell pellet was resuspended in 500 ml of 1 × EM9 salts with gentle shaking, and pelleted again. This step was repeated once more, and the cell pellet was again resuspended in 20 ml of prepared EM9 medium (minus isotopes), and distributed evenly into two 500 ml flasks of EM9 medium. Expression was induced by addition of IPTG (1 mM), and isotopes ($^{15}NH_4Cl$, 1 g l^{-1}) were added as the final step before incubation at 37 °C and 200 rpm.

Incubation proceeded for 10 minutes, at which point rifampicin was added to the flask (final concentration 150 mg/L), and incubation continued for a further 35 minutes. The cells were then immediately chilled on ice, before harvesting at 4000 rpm for 20 minutes. The cells could be stored at this stage at -20 °C, after being frozen in liquid nitrogen.

The expression time is optimal for maximising the extent to which the ribosomes in the culture are occupied with stalled nascent chains and minimising the extent of background isotopic labelling and nascent chain release. The rifampicin is used to inhibit the activity of RNA polymerase [117], preventing further ribosome growth in the labelled medium, and hence minimising the extent of background ribosome labelling.

7.5.1 Growth and expression of ^{15}N-^{19}F labelled nascent chains

The method described in the above section was modified to include a step for the incorporation of fluorine. After the resuspension of cells in the inducing EM9 medium and addition of isotopes, glyphosate (500 mg l^{-1}) was added alongside 3-fluorotyrosine (35 mg l^{-1}), phenylalanine (30 mg l^{-1}) and tryptophan (30 mg l^{-1}). The culture was incubated for 10 minutes before induction by addition of IPTG, at which point the protocol resumed the sequence described in the above section (i.e., addition of rifampicin after 10 minutes, followed by harvesting of cells after a further 35 minutes).

7.5.2 Purification of nascent chains; sucrose gradient method

The cell pellet was thawed in a water bath at 37 °C, and resuspended in lysis buffer. The cells were then lysed by french press (6 passes, 1000 psi), pelleted by centrifugation (18,000 rpm, 45 min, SS3 4 rotor), and the soluble fraction loaded onto a sucrose cushion (30 %). The cushion was centrifuged at 30,000 rpm for 16 hours, yielding a brown pellet at the bottom of the tube. After disposal of the sucrose cushion and the

careful removal of excess sucrose with a clean paper towel, the tube was inverted, to allow the brown pellet to gently slide off, revealing a clear pellet underneath, which corresponds to the ribosome fraction. This pellet was resuspended in buffer B1 (5 ml per litre of growth) with gentle shaking at 4 °C. To accelerate resuspension, a round edged glass rod was used to gently perturb the pellet into the buffer.

The resuspended pellet was bound to an IDA-Nickel resin equilibrated in buffer B1 for 2 hours at 4 °C. After washing in B1 for 150 ml, the nascent chain was eluted in buffer B2, collecting the first 150 ml of eluent (Bradford reagent was used to assay when the sample had eluted).

The eluent was immediately diluted by half in buffer B4, and concentrated to approx. 5 ml using a 100 kDa cutoff centrifugal filter. The total yield and purity of ribosomes was measured at this stage by evaluating absorbance at 260 and 280 nm, and a maximum of 6000 pmol was retained for further purification. Any excess was frozen in liquid nitrogen and stored at -80 °C.

The 6000 pmol of crude sample that was retained was then passed through a 10-30% sucrose gradient. A total of 6 type45Ti tubes were each loaded with ca. 1000 pmol of ribosome solution and centrifuged (22,000 rpm, 15 hours, 4 °C). The gradient was then collected in 1 ml fractions by suction from the bottom of the tube, and fractions were monitored with a 280 nm absorbance readout. The pure 70S containing fractions were identified by silverstained SDS-PAGE (Figure 4.2B), pooled, concentrated and exchanged into Tico buffer (5×, 15 ml to 0.5 ml) before being snap frozen in liquid nitrogen and stored at -80 °C.

Concentration was evaluated by measurement of the OD_{260}. It is known from the literature that the following formula yields the ribosome concentration in μM;

$$[\text{ribosomes}]\,(\text{pmol ml}^{-1}) = OD_{260} \times 24 \tag{7.5.1}$$

To assess purity, the OD_{260}/OD_{280} ratio was measured to be 1.8, which is within the acceptable range for bacterial ribosomes of 1.9 ± 0.1. The yield from a single gradient of six tubes was typically 0.5 ml of 8 μM sample. The typical nascent chain yield from 1 litre of culture was typically 4-6000 pmol.

7.5.3 Purification of nascent chains; butyl column method

Further to the method described above, a more efficient protocol has since been developed for the purification of nascent chains. This protocol, detailed below, was used for the purification of the ^{19}F labelled +31 and +67 FLN750 nascent chains described in Chapter 5.4.

After cell lysis and centrifugation, the supernatant was loaded onto a sucrose cushion and centrifuged for

4.5 hours at 44,000 rpm. After resuspension of the pellet in buffer B2, the solution was bound to a nickel IDA resin and eluted as described above. The eluent was concentrated and loaded onto a second sucrose cushion, and centrifuged for 16 hours at 30,000 rpm. At this stage, the pellet was again resuspended, in buffer $B5_{hs}$ (1.5 M NH_4SO_4, 20 mM Hepes, 400 mM KOAc, 12 mM MgOAc, 2 mM BME, 0.1 % protease inhibitors, pH 7.5).

This sample was then loaded onto a hydrophobic exchange column equilibrated in $B5_{hs}$. The column was washed to baseline in high salt, followed by a further wash step at 80 % of the maximum salt concentration. Ribosomes were eluted by application of a reverse gradient of high to low salt concentration (buffer $B5_{hs}$ to buffer $B5_{ls}$. Buffer $B5_{ls}$ is identical to buffer $B5_{hs}$ with no ammonium chloride). The 70S containing fractions of the eluent were pooled, concentrated, flash frozen in liquid nitrogen, and stored at -80 °C.

7.6 Production and purification of unlabelled 70S ribosomes

The following standard solutions of salts and trace metals were used;
50×salts solution: 1.25 M Na_2HPO_4, 1.25 M KH_2PO_4, 2.5 M NH_4Cl, 0.25 M Na_2SO_3, filtered.

1000×trace metals solution: 50 mM $FeCl_2$ [dissolved in 0.1 M HCl], 20 mM $CaCl_2$, 1 mM each of {$MnCl_2.4H_2O$, $ZnSO_4.7H_2O$}, 2 mM each of {$CoCl_2.6H_2O$, $CuCl_2.2H_2O$, $NiCl_2.6H_2O$, $Na_2MoO_2.2H_2O$, $Na_2SeO_3.5H_2O$, H_3BO_3}, all filtered.

BL21 DE3 trigger factor knockout (ΔTF) *E. coli* cells were cultured for 16 hours on LB agar, before inoculation of a single colony into MDG medium (1×salts, 0.2% (v/v) L-aspartic acid, 0.4%(v/v) glucose, 0.2×trace metals, 2 mM $MgSO_4$, 5 ml) and aeration for 16 hours (37 °C, 250 rpm). 500 µl of this mixture was added to 1 litre of MDG medium before further aeration for 22 hours (OD_600 = 4-6). The culture was harvested (4000 rpm, 20 minutes, Beckmann JLA 8100) and the cell pellet was stored at -20 °C.

The cells were resuspended in lysis buffer (50 mM Hepes, 1 M KOAc, 12 mM $Mg(OAc)_2$, 2 mM β-mercaptoethanol, 5 mM EDTA, 0.2% w/v lysozyme, trace DNASE, protease inhibitor (Roche, 1 tablet per 50 ml), pH 7.5, 15 ml), and lysed by French press (∼1000 psi, 3 passes). After the third pass, the sample acquired a darker brown hue and was notably less viscous. The sample was pelleted (18,000 rpm, 45 min, SS34).

The supernatant (15 ml) was divided amongst 6 sucrose cushions (45 ml, 30% in Tico buffer) in type45Ti

tubes, and centrifuged (40,000 rpm, 12 hours, 4 °C). After wiping away excess sucrose, the clear pellets were resuspended in a total of 10 ml of buffer B (50 mM Hepes, 1 M KOAc, 12 mM $Mg(OAc)_2$, 2 mM β-mercaptoethanol, 5 mM EDTA, 5 mM ATP, pH 7.50). Concentration was estimated by OD_{260} measurements to be 4 μM in 10 ml.

The crude ribosome samples were passed through a 10-30% sucrose gradient. A total of 6 type45Ti tubes were each loaded with ca. 1000 pmol of ribosome solution and centrifuged (22,000 rpm, 15 hours, 4 °C). The fractions containing 70S ribosomes were collected and pooled together before exchanging into Tico buffer (plus protease inhibitors). Concentration was evaluated by measurement of the OD_{260}. To assess purity, the OD_{260}/OD_{280} ratio was measured to be 1.8, which is within the acceptable range for bacterial ribosomes of 1.9 ± 0.1. Silverstained SDS-PAGE also confirmed the purity of the ribosome samples obtained. The yield from a single gradient of six tubes was typically 0.5 ml of 8 μM sample.

7.6.1 Production of ^{19}F labelled 70S ribosomes

Fluorine labelled ribosomes were produced via two approaches - the glyphosate approach and the auxotrophic strain approach;

Glyphosate approach

BL21 DE3 ΔTF *E. coli* cells were grown in 500 ml of MDG medium to a cell density of 0.54. The culture was supplemented with glyphosate (500 mg l^{-1}, L-phenylalanine (80 mg l^{-1}), L-tryptophan (80 mg l^{-1}) and D/L-3-fluorotyrosine (70 mg l^{-1}). The culture was then incubated at 37 °C and 230 rpm for 16 hours (OD_{600} = 3.7), and harvested by centrifugation (4000 rpm, 20 minutes).

Auxotroph approach

Untransformed C43 (DE3) ML14 *E. coli* cells were grown for 16 hours (37 °C 200 rpm, OD_{600} = 1.47) in 500 ml of MDG medium supplemented with tyrosine (7.2 mg l^{-1}) and 3-fluorotyrosine (72 mg l^{-1}). Cells were then washed and resuspended in EM9 medium with the same aromatic supplement. Incubation proceeded at 200 rpm and 37 °C for 10 minutes, followed by addition of 1 mM IPTG, and then rifampicin after another 10 minutes. After rifampicin was added, the culture was incubated for 35 minutes at 37 °C before harvesting. This sample was produced under conditions that replicate the expression of a nascent chain construct, with the aim of acting as a control for background labelling of ribosomes that may occur from expression of nascent chains. Purification of the fluorinated ribosomes then proceeded as described in section 7.6.

7.7 Preparation of sucrose gradients

Two stock solutions of sucrose were prepared in buffer B4, at 10 % and 30 % (w/v) sucrose. To prepare the gradient, 17 ml of each solution was poured into a gradient maker, which mixed the solutions together as they were drawn by suction into the SW28 tube. Mixing began with mostly 30 % sucrose, and gradually became more dominated by the lower concentration as the tube filled up. This gave 34 ml of a linear 30-10 % sucrose concentration gradient.

7.8 Immunoblotting with anti-his and anti-SecM antibodies

After electrophoresis, the contents of the acrylamide gel was wet transferred onto a nitrocellulose blotting membrane in the presence of a transfer buffer (20 % MeOH, 14.4 g l^{-1} glycine, 6.6 g l^{-1} Tris-Base, 1 % SDS). Transfer was induced at 25 volts, 250 milli amps and 25 watts for 2 hours.

For anti-his immunoblotting, the membrane was blocked by incubation at room temperature in a 1 % (v/v) casein solution in TBS buffer (2.42 g l^{-1} Tris-Base, 11.69 g l^{-1} NaCl, pH 7.4) for 1 hour. The membrane was then incubated in 5 ml of a 1/5000 dilution of anti-his antibody (Qiagen) in TBS-T buffer (TBS, 0.05 % tween-20) at 4 °C for 2 at least hours, or overnight. The membrane was washed in TBS-T (3 × 10 minutes) and visualisation was induced by addition of 1 ml of a peroxide solution followed by 1 ml of enhancer (SuperSignal West Pico Chemiluminescent Substrate Kit, Thermoscientific) and the membrane was visualised by detection of chemiluminescence.

For anti-SecM immunublotting, membrane blocking was carried out in 5 % BSA in TBS-T for 1 hour at room temperature. The membrane was then incubated in a 1/10,000 dilution of the primary antibody (anti-SecM) in 2 % BSA and TBS-T at 4 °C for at least 2 hours. The membrane was washed in TBS-T (4 × 10 minutes) and bound to a 1/20,000 dilution of the secondary antibody (anti-rabbit, Cell Signalling Technology) in 2 % BSA and TBS-T. The membrane was washed again in TBS-T (4 × 10 minutes) and visualised with a combination of the pico (1 ml each of peroxide and enhancer) and femto (100 µl) visualisation solutions.

7.9 NMR experimental data collection

NMR spectra were collected on Bruker instruments at 500, 600, 700 and 800 MHz as described in the main body of the text. All data was processed in NMRpipe, and further analysis and data fitting procedures were

Wild type α-syn titration		His-tagged α-syn titration	
[α-syn] / μM	[ribosomes] / μM	[H_6-αsyn] μM	[ribosomes] / μM
8.9	9.5	6.7	6.4
19.5	9.3	18.5	6.1
39.4	8.9	46.3	5.6
90.0	7.9	103.4	5.3
170.9	7.5	223.3	4.6
380.1	6.4	317.5	4.0
641.5	5.3	-	-

Table 7.1: *Concentrations used in the α-syn and H_6-αsyn ribosome titrations.* All samples prepared from stock solutions of ribosomes, α-syn and H_6-αsyn as required. All samples consisted of Tico buffer in 10 % D_2O and 0.01 % DSS, with 0.1 × protease inhibitor cocktail.

carried out in matlab.

7.9.1 Titration of α-syn and H_6-αsyn with ribosomes

All NMR samples were prepared by exchanging α-syn, H_6-αsyn, or ribosome solution into Tico buffer (pH 7.5) containing the standard Roche protease inhibitor cocktail (1 tablet for each 100 ml), and 10% D_2O.

For both the α-syn and H_6-αsyn titrations, the stock solutions of protein and ribosomes were mixed in various combinations to produce a range of concentration ratios, which are detailed in Table 7.1.

7.9.2 Urea titrations of isolated proteins

Each urea titration was carried out by one of two approaches, as detailed below;

Approach one - Two stock solutions of the protein of interest were prepared with equivalent protein concentrations. One stock solution contained no urea, and the second stock was high in urea concentration (10 M). The titration was then constructed by dilution of the urea to appropriate concentrations by mixing appropriate volumes of the two samples. This approach was used for the HP36 titration.

Approach two - Alternately, urea was added directly to the protein sample as a solid. the change in volume was accounted for by taking the molal volume of urea and using this to calculate the change in volume upon addition of a given mass of urea to an aqueous solution of known volume at 25 °C. Calculations were made using data obtained from Motin et al. [151]. This approach was used for the GA module titration. All titrations were carried out at pH 5.0.

7.10 Circular dichroism experiments

Circular dichroism data was collected on a chirascan plus CD spectrometer equipped with a xenon lamp. Before switching on the lamp, the system was purged with a nitrogen atmosphere.

CD spectra were acquired by measuring the circular dichroism of a sample in 1 nm increments, wth 1 second for each measurement. For temperature melts, CD spectra were acquired at increments of 1 °C between 25 and 90 °C. Urea titrations were prepared in the manner as described above in section 7.9.2.

7.11 Data treatment

Raw NMR data was processed using NMRpipe [152], and diffusion coefficients were determined using the NMRpipe package dosyView. The ccpn analysis tool [119] was used to visualise HSQC spectra and was also used to measure peak intensities. The curve fitting procedures used to determine binding constants from the relaxation experiments data were executed using MATLAB [67].

7.12 Error analysis

Error propagation was carried out using standard methods, as described by Bevington and Robinson [153]. For the combination and transformation of the variables u and v, the resulting errors are described below, where a and b are constants.

$$x = u \pm a \qquad \sigma_x = \sigma_u \tag{7.12.1a}$$

$$x = au \pm bv \qquad \sigma_x^2 = a^2\sigma_u^2 + b^2\sigma_v^2 \pm ab\sigma_{uv}^2 \tag{7.12.1b}$$

$$x = auv \qquad \sigma_x^2 = \left(\frac{\sigma_u^2}{u^2} + \frac{\sigma_v^2}{v^2} + \frac{2\sigma_{uv}^2}{uv}\right) \tag{7.12.1c}$$

$$x = a\frac{u}{v} \qquad \sigma_x^2 = \left(\frac{\sigma_u^2}{u^2} + \frac{\sigma_v^2}{v^2} - \frac{2\sigma_{uv}^2}{uv}\right) \tag{7.12.1d}$$

$$x = au^b \qquad \sigma_x = bx\frac{\sigma_u}{u} \tag{7.12.1e}$$

$$x = e^{bu} \qquad \sigma_x = bx\sigma_u \tag{7.12.1f}$$

$$x = a^{bu} \qquad \sigma_x = (bx\ln a)\sigma_u \tag{7.12.1g}$$

$$x = a\ln bu \qquad \sigma_x = a\frac{\sigma_u}{u} \tag{7.12.1h}$$

The σ_{uv} term is a cross correlation error, which is 0 when u and v are uncorrelated.

Appendix A

NMR Lineshape Simulation Script

The following Matlab script was used to produce the lineshape simulations presented in Chapter 2.

```
tic                    % Set timer
clear all              % Always ensure the workspace is empty
kex = [10 100000].*2*pi; % Set slow and fast exchange kinetics. Multiply by 2*pi for comparison
    with angular frequency
pA = 0:0.1:1;          % set unfolded state populations
pB = 1-pA;             % set folded state populations
% For the purposes of comparison, pB is equivalent to pF, and pA is
% euivalent to pU



% Set appropriate relaxation rates for amide protons and methyl protons in
% folded and unfolded protein states.
R2A_15N = 25*2*pi;    % estimated unfolded peak relaxation rate (amide proton)
R2B_15N = 300*2*pi;   % estimated folded peak relaxation rate (amide proton)
R2A_13C = 25*2*pi;    % estimated unfolded peak relaxation rate (methyl proton)
R2B_13C = 53*2*pi;    % estimated folded peak relaxation rate (methyl proton)

% Set realistic values for chemical shifts of folded and unfolded states
wA = 350*2*pi;
wB = 0*2*pi;
```

```
% NMR experiment acquisition parameters. These represent real experimental
% parameters that would be set up for the acquisition of a 1D experiment
sample_time = 1/5000;               % Sampling frequency of experiment in seconds (i.e., 0.2 ms per
     signal measurement)
t = 0:sample_time:(1023 *sample_time); % points in time at which data is collected
at = t(end);                        % total amount of time in which the FID signal is collected
sw = 1/sample_time;                 % sweep width (defines maximum resolvable frequency. Also the
    offset)
res = 1/at;                         % resolution of signal in Hz
S2N = [5 10 15 20 25 30 35 40];     % The S2N variable is used to control the amount of artifical
    noise added to the simulated signal

% Generate frequency domain values for after fourier transformation of the time domain signal
freq = -sw/2:res:sw/2;

% Optional Apodisation window function. Simulations also works well without this function
apod = cos(pi/2*t/max(t)).^2;

% Generate data for all values of S2N, kex and pB
for k = 1:length(S2N)
   for n = 1:length(kex)
       for o = 1:length(pB)
           for q = 1:20      % Repeat 20 times so that error in frequency can be measured for fast
               exchange spectra
               Lambda_13C =  [-(R2A_13C-1i*wA)-pB(o).*kex(n)      kex(n).*(1-pB(o))
                           pB(o).*kex(n)         (-(R2B_13C-1i*wB) + kex(n).*(pB(o) - 1)) ];
               Lambda_15N =  [-(R2A_15N-1i*wA)-pB(o).*kex(n)      kex(n).*(1-pB(o))
                           pB(o).*kex(n)         (-(R2B_15N-1i*wB) + kex(n).*(pB(o) - 1)) ];
               M0 = [pA(o)
                     pB(o)];
                  for p = 1:length(t)
                      M_13C = expm(Lambda_13C*t(p))*M0;
                      M_15N = expm(Lambda_15N*t(p))*M0;
                      FID_13C(p,k,n,o,q) = M_13C(1)+M_13C(2);
```

```
                    FID_15N(p,k,n,o,q) = M_15N(1)+M_15N(2);
                end
            % add noise
            FID_13C(:,k,n,o,q) = awgn(FID_13C(:,k,n,o,q),S2N(k));
            FID_15N(:,k,n,o,q) = awgn(FID_15N(:,k,n,o,q),S2N(k));
            % apply window function
            FID_13C(:,k,n,o,q) = FID_13C(:,k,n,o,q) .* apod';
            FID_15N(:,k,n,o,q) = FID_15N(:,k,n,o,q) .* apod';
        end
    end
end

% The FID_13C and FID_15N variables now comprise 5D matrices containing FID
% signal generated for a range of values of exchnage kinetics (kex), A-B
% population distribuitons (pB) and signal to noise (S2N). For each value
% of kex, pB and S2N, we therefore have a time domain signal corresponding
% to those parameters.
% For each value of kex, pB and S2N, we also
% generated the same signal 20 times, as the stochastic nature of the noise
% leads to an error in peak position which can be estimated through
% repetition.

% We now take the Fourier transform the FID signal generated above
for k = 1:length(S2N)
    for n = 1:length(kex)
        for o = 1:length(pB)
            for q = 1:20
                FT_13C(:,k,n,o,q) = (fftshift(fft(FID_13C(:,k,n,o,q))));
                FT_15N(:,k,n,o,q) = (fftshift(fft(FID_15N(:,k,n,o,q))));
            end
        end
    end
end
```

```
toc    % Total time for simulation was 7 hours (macbook pro late 2013 2.6GHz intel i7 processor
    with 16GB RAM)
save simulation_data % Allow for the quick loading of the simulated dataset for future testing
%%
 load simulation_data % Remove comment to load simulated dataset if you
% already have it


%% Plot spectra
% This section of script reproduces the spectra displayed in figures 2.1,
% 2.2, 2.3 and 2.4 in chapter 2.
%% Plot methyl proton spectra, slow exchange
for p = 1                  % Select slow exchange spectra
   for k = 1:length(S2N)   % All signal to noise values
       for o = 1:length(pB) % All values of pB
           subplot(2,4,9-k) % Plot spectra with the strongest signal to noise first
           for q = 1
               hold on
               test13C = reshape(real(FT_13C(410:666,k,p,:,q)),length(freq(410:666)),length(pB))'; %
                    Extract desired component of matrix and reshape/transpose.
               w = waterfall(freq(410:666),pB,test13C);
               set(w,'edgecolor','k','linewidth',1.5)
               set(gca,'fontsize',30)
               highlight13C = reshape(real(FT_13C(481:544,k,p,:,q)),length(freq(481:544)),length(pB)
                   )'; % Plot in red the peak we intend to use for measurement of pB/pF.
               h = waterfall(freq(481:544),pB,highlight13C);
               set(h,'edgecolor','r','linewidth',1.5)
               xlim([-500 750])
               ylim([0 1])
               zlim([-5 30])
               view(5, 50)
%                xlabel('Frequency / s^{-1}')
%                ylabel('p_F')
%                zlabel('Signal')
           end
```

```
        end
    end
end

%% Plot methyl proton spectra, fast exchange
for p = 2                % Select fast exchange spectra
    for k = 1:length(S2N)
        for o = 1:length(pB)
            subplot(2,4,9-k)
            for q = 1
                hold on
                test13C = reshape(real(FT_13C(410:666,k,p,:,q)),length(freq(410:666)),length(pB))'; %
                     Extract desired component of matrix and reshape/transpose.
                w = waterfall(freq(410:666),pB,test13C);
                set(w,'edgecolor','k','linewidth',1.5)
                set(gca,'fontsize',30)
                xlim([-500 750])
                ylim([0 1])
                zlim([-5 30])
                view(5, 50)
                %xlabel('Frequency / s^{-1}')
                %ylabel('p_F')
                %zlabel('Signal')
            end
        end
    end
end
%% Plot amide proton spectra, slow exchange
for p = 1
    for k = 1:length(S2N)
        for o = 1:length(pB)
            subplot(2,4,9-k)
            for q = 1
                hold on
```

```
            test15N = reshape(real(FT_15N(410:666,k,p,:,q)),length(freq(410:666)),length(pB))'; %
                 Extract desired component of matrix and reshape/transpose.
            w = waterfall(freq(410:666),pB,test15N);
            set(w,'edgecolor','k','linewidth',1.5)
            set(gca,'fontsize',30)
            highlight15N = reshape(real(FT_15N(554:607,k,p,:,q)),length(freq(554:607)),length(pB)
                )'; % Plot in red the peak we intend to use for measurement of pB/pF.
            h = waterfall(freq(554:607),pB,highlight15N);
            set(h,'edgecolor','r','linewidth',1.5)
            xlim([-500 750])
            ylim([0 1])
            zlim([-5 30])
            view(5, 50)
%             xlabel('Frequency / s^{-1}')
%             ylabel('p_B')
%             zlabel('Signal')
        end
    end
end

%% Plot amide proton spectra, fast exchange
for p = 2 % exchange rate
    for k = 1:length(S2N)
        for o = 1:length(pB)
            subplot(2,4,9-k)
            for q = 1 % experiment repeats
                hold on
                test15N = reshape(real(FT_15N(410:666,k,p,:,q)),length(freq(410:666)),length(pB))'; %
                     Extract desired component of matrix and reshape/transpose.
                w = waterfall(freq(410:666),pB,test15N);
                set(w,'edgecolor','k','linewidth',1.5)
                set(gca,'fontsize',30)
                xlim([-500 750])
```

```
                ylim([0 1])
                zlim([-5 30])
                view(5, 50)
%                 xlabel('Frequency / s^{-1}')
%                 ylabel('p_B')
%                 zlabel('Signal')
            end
        end
    end
end

% After confirming the spectra, we now proceed to calculate populations from these spectra using
% measurments of peak heights (slow exchange) or peak position (fast
% exchange)
%% Measurement of peak heights for slow exchange datasets
% For amide protons we measure the height of the sharper, unfolded peak at
% ca 350 Hz. Inspection of the freq variable shows that this is point 585 in
% the frequency domain signal.

% For methyl protons, we measure the height of the broader, folded peak at
% ca. 0 Hz. Inspection of the freq variable shows that this is point 513 in
% the frequency domain signal.

for p = 1 % Investigate slow exchange spectra
    for k = 1:length(S2N)
        for o = 1:length(pB)
            for q = 1 % No need to repeat experiments for peak heights. Error comes from noise
                height_13C(o,k) = (real(FT_13C(513,k,p,o,q))); % Height at point 513 (2.4438 Hz)
                height_15N(o,k) = (real(FT_15N(585,k,p,o,q))); % Height at point 585 (354.35 Hz)
            end
        end
    end
end
```

```
% measure noise in spectra to determine error in peak height
for p = 1
    for k = 1:length(S2N)
        for o = 1:length(pB)
            for q = 1
                Sigma_13C_height(o,k) = std(real(FT_13C(881:1024,k,p,o,q))); % Between point 881 and
                    1024 we only see noise
                Sigma_15N_height(o,k) = std(real(FT_15N(881:1024,k,p,o,q)));
            end
        end
    end
end

% Calculate pB values
for k = 1:length(pB)
    pB_calc_13C_slow(k,:) = height_13C(k,:)./(height_13C(1,:)); % Normalise with respect to highest
         peak
    pA_calc_15N_slow(k,:) = height_15N(k,:)./(height_15N(end,:));
end
% propagate errors in calculated pB values (see Error analysis section in methods for details of
     error propagation)
for k = 1:length(pB)
    pB_err_13C_slow(k,:) = sqrt(((Sigma_13C_height(k,:).^2)./(height_13C(k,:)).^2) + ((
        Sigma_13C_height(1,:).^2)./(height_13C(1,:))));
    pA_err_15N_slow(k,:) = sqrt(((Sigma_15N_height(k,:).^2)./(height_15N(k,:)).^2) + ((
        Sigma_15N_height(end,:).^2)./(height_15N(end,:))));
end

%% Regression analysis of slow exchange datasets
for k = 1:length(S2N)
    [A] = fitlm(1-pA_calc_15N_slow(:,k),pB);
    residual_15N_slow(k) = A.Rsquared.Ordinary; % R-square values for amide proton slow exchange
        dataset
    [B] = fitlm(pB_calc_13C_slow(:,k),pB);
```

```
    residual_13C_slow(k) = B.Rsquared.Ordinary; % R-square values for methyl proton slow exchange
        dataset
end
%% Plot calculated versus true pF values for methyl proton spectra in slow exchange
figure
for k = 1:length(S2N)
    for q = 1
        subplot(2,4,9-k)
        hold on
        errorbar(pB,pB_calc_13C_slow(:,k),pB_err_13C_slow(:,k),'kx','linewidth',1.5)
        plot(0:1,0:1,'k','linewidth',1.5)
        set(gca,'fontsize',30)
        xlabel('p_F')
        ylabel('p_{F,fit}')
        xlim([-0.03 1.03])
        ylim([-0.03 1.03])
        title(num2str(residual_13C_slow(k),'%.3f'),'fontsize',15) % show Rsquared value on plots
        box on
    end
end
%% Plot calculated versus true pF values for amide proton spectra in slow exchange
figure
for k = 1:length(S2N)
    for q = 1
        subplot(2,4,9-k)
        hold on
        errorbar(pB,1-pA_calc_15N_slow(:,k),pA_err_15N_slow(:,k),'kx','linewidth',1.5)
        plot(0:1,0:1,'k','linewidth',1.5)
        set(gca,'fontsize',30)
        xlabel('p_F')
        ylabel('p_{F,fit}')
        xlim([-0.03 1.03])
        ylim([-0.03 1.03])
        title(num2str(residual_15N_slow(k),'%.3f'),'fontsize',15) % show Rsquared value on plots
```

```
        box on
    end
end
%% Measure chemical shifts for fast exchange datasets
for p = 2 % fast exchange
    for k = 1:length(S2N)
        for o = 1:length(pB)
            for q = 1:20 % We now repeat each experiemnt 20 times, so that variation in peak
                position as a result of the addition of noise can be evaluated.
                clear max_13C position_13C max_15N position_15N
                [max_13C, position_13C] = max(real(FT_13C(:,k,p,o,q)));
                omega_13C(o,q,k) = freq(position_13C)-freq(513); % Offset of approx 2.44. Eliminate
                    this so that initial frequency is 0.
                [max_15N, position_15N] = max(real(FT_15N(:,k,p,o,q)));
                omega_15N(o,q,k) = freq(position_15N)-freq(513); % Offset of approx 2.44. Eliminate
                    this so that initial frequency is 0.
                clear max_13C position_13C max_15N position_15N
            end
        end
    end
end
% Take average chemical shift for each experiment, and measure standard
% deviation to determine error
for o = 1:length(pB)
    for k = 1:length(S2N)
        omega_13C_avg(o,k) = mean(omega_13C(o,:,k));
        omega_13C_err(o,k) = std(omega_13C(o,:,k));
        omega_15N_avg(o,k) = mean(omega_15N(o,:,k));
        omega_15N_err(o,k) = std(omega_15N(o,:,k));
    end
end
% Minimum error is resolution
omega_13C_err(omega_13C_err <= res) = res;
omega_15N_err(omega_15N_err <= res) = res;
```

```
%% Calculate pB values and propagate errors.

for o = 1:length(pB)
    for k = 1:length(S2N)
        pB_13C_fast(o,:,k) = (omega_13C(o,:,k) - omega_13C(end,1,k))./(omega_13C(1,1,k) - omega_13C(
            end,1,k));
        pB_13C_fast_avg(o,k) = mean(pB_13C_fast(o,:,k));
        pB_13C_fast_err(o,k) = std(pB_13C_fast(o,:,k));
        %
        pB_15N_fast(o,:,k) = (omega_15N(o,:,k) - omega_15N(end,1,k))./(omega_15N(1,1,k) - omega_15N(
            end,1,k));
        pB_15N_fast_avg(o,k) = mean(pB_15N_fast(o,:,k));
        pB_15N_fast_err(o,k) = std(pB_15N_fast(o,:,k));
    end
end
% Define minimum error in pB relative to resolution of spectrum
pB_min_err = 1 - ((res - 350)/(0 - 350));
pB_13C_fast_err(pB_13C_fast_err <= pB_min_err) = pB_min_err;
pB_15N_fast_err(pB_15N_fast_err <= pB_min_err) = pB_min_err;

%% Regression analysis of fast exchange datasets
for k = 1:length(S2N)
    for q = 20;
        [A1] = fitlm(pB,pB_15N_fast(:,q,k));
        residual_15N_fast(k) = A1.Rsquared.Ordinary; % R-square values for amide proton fast
            exchange dataset
        [B1] = fitlm(pB,pB_13C_fast(:,q,k));
        residual_13C_fast(k) = B1.Rsquared.Ordinary; % R-square values for methyl proton fast
            exchange dataset
    end
end
%% Plot calculated versus true pF values for methyl proton spectra in fast exchange
figure
```

```
for k = 1:length(S2N)
    for q = 20
        subplot(2,4,9-k)
        hold on
        errorbar(pB,pB_13C_fast(:,q,k),pB_13C_fast_err(:,k),'kx','linewidth',1.5)
        plot(0:1,0:1,'k','linewidth',1.5)
        set(gca,'fontsize',30)
        xlabel('p_F')
        ylabel('p_{F,fit}')
        xlim([-0.03 1.03])
        ylim([-0.03 1.03])
        title(num2str(residual_13C_fast(k),'%.3f'),'fontsize',15) % show Rsquared value on plots
        box on
    end
end
%% Plot calculated versus true pF values for amide proton spectra in fast exchange
figure
for k = 1:length(S2N)
    for q = 20
        subplot(2,4,9-k)
        hold on
        errorbar(pB,pB_15N_fast(:,q,k),pB_15N_fast_err(:,k),'kx','linewidth',1.5)
        plot(0:1,0:1,'k','linewidth',1.5)
        set(gca,'fontsize',30)
        xlabel('p_F')
        ylabel('p_{F,fit}')
        xlim([-0.03 1.03])
        ylim([-0.03 1.03])
        title(num2str(residual_15N_fast(k),'%.3f'),'fontsize',15) % show Rsquared value on plots
        box on
    end
end
```

Appendix B

NMR Triple Resonance Assignment of the GA Module

B.1 Triple resonance experiments

The triple resonance experiments used to obtain the assignment of the GA module NMR resonances are briefly described here, followed by the the list of resonances. For magnetisation transfer pathways, see Figure B.1, below.

HNCO

This experiment starts with excitation of a backbone amide proton, and transfers magnetisation to the corresponding nitrogen, and then the carbonyl carbon of the same amino acid, before transferring magnetisation back through the nitrogen to be observed on the proton. This is one of the most sensitive triple resonance experiments, and allows observation of the coupling between $^{15}N_H$, $^{1}H_N$ of one residue with the adjacent $^{13}C'$ of the previous residue.

HNCACB

The HNCACB begins with excitation of the $^{1}H_\alpha$ and $^{1}H_\beta$ of the side chain residues i and i-1. This is then transferred through the $^{13}C_\alpha$ and $^{13}C_\beta$ nuclei, followed by ^{15}N of residue i, and finally the $^{1}H_N$ for detection. The C_α and C_β appear in one dimension, and the ^{15}N and the $^{1}H_N$ appear in their own respective

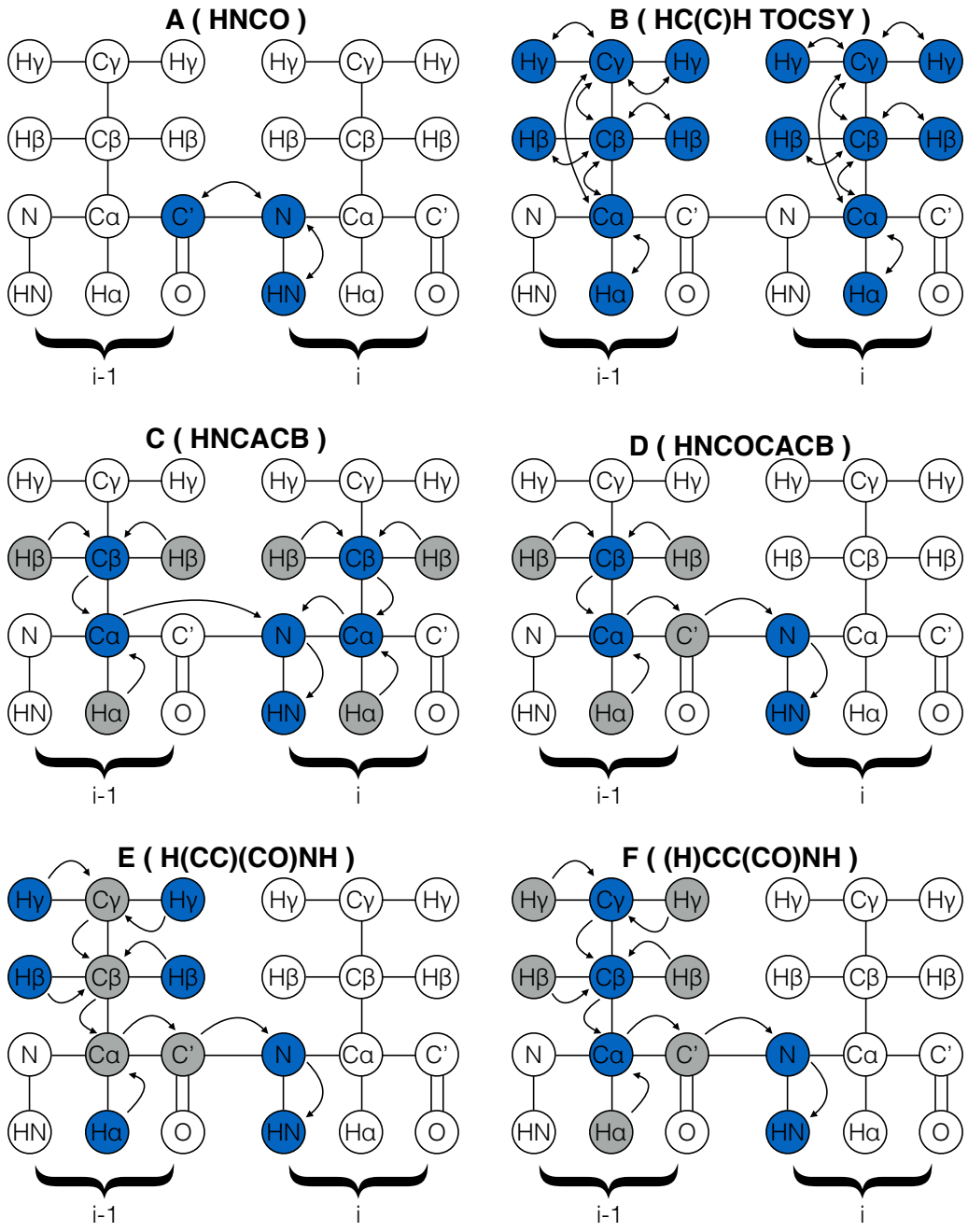

Figure B.1: *Magnetisation transfer pathways for the triple resonance spectra used for assignment in this report.* **A:** HNCO. **B:** HC(C)H TOCSY. **C:** HNCACB. **D:** HNCOCACB. **E:** H(CC)(CO)NH. **F:** (H)CC(CO)NH.
Blue corresponds to nuclei that are detected in the final experiment. Grey corresponds to nuclei that are either excited initially with a radio-frequency pulse, or used to transfer magnetisation, without being observed in the final experiment. This figure was produced using the Keynote software. General design of the diagrams was obtained from [63].

dimensions. Overall, this experiment allows observation of coupling between the amide of residue i and the C_α and C_β nuclei of residues i and i-1.

HNCOCACB

Almost identical to the HNCACB, however an extra step includes transfer from C_α through the adjacent carbonyl carbon to the ^{15}N of the adjacent residue. Hence we exclusively observe a correlation between the amide of residue i and the C_α and C_β of i-1. This can be combined with the HNCACB to connect adjacent residues.

H(CC)(CO)NH

This experiment begins with excitation of aliphatic side chain ^1H nuclei (i-1), and transfers magnetisation through all aliphatic ^{13}C nuclei, including the carbonyl, to the ^{15}N and ^1H of residue i. Magnetisation is evolved for detection on the aliphatic protons, and the nitrogen and amide proton.

(H)CC(CO)NH

This is very similar to the H(CC)(CO)NH experiment, however it is the aliphatic ^{13}C nuclei of the side chains of the preceding residue that are observed in this case, as opposed to the aliphatic ^1H nuclei. These two experiments allow the observation of coupling between the amide group of residue i to the aliphatic side chain ^1H and ^{13}C nuclei of residue i-1.

HC(C)H TOCSY

The TOCSY experiment begins with excitation of the aliphatic ^1H nuclei of a given residue. This is then transferred to the ^{13}C nuclei, where an isotropic mixing period allows couplings between each proton and all other spins in the system to evolve. Magnetisation is then transferred back to ^1H for detection [154].

The TOCSY experiment allows for side-chain assignment, which can be used to easily identify an amino acid for a given C_α frequency. As a result of this, the connectivities between amino acids as determined from HNCACB and others can be assigned a location in the primary sequence, leading to residue specific assignment of the spectra.

With the above experiments, we can produce a series of 3D spectra that allow us to show which resonances are connected to each other, and with some identification using TOCSY and chemical shifts, along with the

known primary sequence of the protein, the identity and connectivity of these resonances can be verified in the 2D HSQC spectra of the protein.

B.2 GA module resonance assignment

Shown below is the amide backbone and carbon backbone/sidechain assignment of the GA module of protein PAB as discussed in this thesis. Chemical shifts are given in ppm.

Residue number	Amino acid	H_N	N_H	C'	Cα	Cβ
1	Thr	-	-	176.05	62.08	70.41
2	Ile	8.52	122.38	-	62.86	38.16
3	Asp	8.24	121.19	178.69	56.79	40.90
4	Gln	8.13	120.68	181.04	58.68	28.42
5	Trp	-	-	178.51	58.71	28.39
6	Leu	8.29	123.83	177.78	60.58	28.81
7	Leu	8.53	119.55	180.48	58.04	41.82
8	Lys	7.86	120.10	179.15	59.73	32.25
9	Asn	8.05	117.36	176.71	55.99	37.95
10	Ala	8.08	121.75	181.28	55.14	18.50
11	Lys	8.24	119.62	177.58	60.65	33.16
12	Glu	7.93	117.82	180.31	59.70	29.23
13	Asp	8.62	120.81	178.42	57.28	40.32
14	Ala	8.12	122.67	179.40	55.45	18.88
15	Ile	8.31	118.54	177.54	65.92	37.70
16	Ala	7.92	121.22	181.05	55.54	17.89
17	Glu	8.15	120.01	179.82	59.84	29.72
18	Leu	8.40	121.07	179.28	58.83	40.62
19	Lys	8.61	119.38	181.91	60.36	32.65
20	Lys	8.13	121.93	177.57	59.29	32.33
21	Ala	7.47	120.06	177.15	52.14	18.88

22	Gly	7.88	106.26	174.12	45.31	-
23	Ile	8.04	123.84	175.34	61.26	37.29
24	Thr	8.21	114.06	-	60.92	70.30
25	Ser	6.95	116.45	-	57.91	63.57
26	Asp	-	-	-	-	-
27	Phe	-	-	178.46	60.98	39.01
28	Tyr	7.11	117.45	177.46	59.76	37.74
29	Phe	7.36	119.25	178.06	58.49	36.45
30	Asn	8.69	117.36	177.06	56.15	38.02
31	Ala	7.40	121.69	180.32	55.35	17.60
32	Ile	7.73	117.86	177.76	63.83	36.78
33	Asn	8.18	115.53	177.90	55.94	38.47
34	Lys	7.69	118.88	176.80	56.40	33.00
35	Ala	7.38	124.28	177.86	53.32	19.96
36	Lys	8.74	116.98	176.78	57.38	34.90
37	Thr	7.22	105.31	174.73	58.82	73.39
38	Val	9.11	122.89	177.41	67.08	31.72
39	Glu	8.90	119.58	179.57	60.59	28.63
40	Glu	8.05	121.68	178.70	59.37	30.70
41	Val	7.91	119.42	177.14	67.61	31.51
42	Asn	8.20	116.19	177.25	56.36	38.21
43	Ala	8.21	122.84	180.99	55.32	18.16
44	Leu	8.32	120.28	178.57	57.90	42.87
45	Lys	8.63	118.53	177.60	60.41	31.44
46	Asn	8.07	115.25	177.85	56.45	38.24
47	Glu	8.07	121.31	179.31	59.54	29.74
48	Ile	8.43	121.35	178.66	65.19	38.22
49	Leu	8.28	119.62	180.24	58.22	41.56
50	Lys	7.89	118.80	-	58.91	32.33
51	Ala	7.71	120.34	178.17	53.58	18.52
52	His	7.63	116.27	173.78	56.97	30.45

53	Ala	7.49	129.91	-	54.30	20.00

References

[1] L. D. Cabrita, S.-T. D. Hsu, H. Launay, C. M. Dobson, and J. Christodoulou. Probing ribosome-nascent chain complexes produced in vivo by NMR spectroscopy. *Proc. Natl. Acad. Sci. USA*, 106(52):22239–44, December 2009.

[2] C. M. Kaiser, D. H. Goldman, J. D. Chodera, I. Tinoco, and C. Bustamante. The ribosome modulates nascent protein folding. *Science*, 334(6063):1723–7, December 2011.

[3] A. M. Knight, P. H. Culviner, N. Kurt-Yilmaz, T. Zou, S. B. Ozkan, and S. Cavagnero. Electrostatic Effect of the Ribosomal Surface on Nascent Polypeptide Dynamics. *ACS Chem. Biol.*, 8(6):1195–1204, 2013.

[4] A. N. Fedorov and T. O. Baldwin. Contribution of cotranslational folding to the rate of formation of native protein structure. *Proc. Natl. Acad. Sci. USA*, 92(4):1227–31, February 1995.

[5] T. Tsalkova, O. W. Odom, G. Kramer, and B. Hardesty. Different conformations of nascent peptides on ribosomes. *J. Mol. Biol.*, 278(4):713–23, May 1998.

[6] J. P. Ellis, C. K. Bakke, R. N. Kirchdoerfer, L. M. Jungbauer, and S. Cavagnero. Chain dynamics of nascent polypeptides emerging from the ribosome. *ACS Chem. Biol.*., 3(9):555–66, October 2008.

[7] B. Seidelt, C. A. Innis, D. N. Wilson, M. Gartmann, J. P. Armache, E. Villa, L. G. Trabuco, T. Becker, T. Mielke, K. Schulten, T. A. Steitz, and R. Beckmann. Structural insight into nascent polypeptide chain-mediated translational stalling. *Science*, 326(5958):1412–5, December 2009.

[8] C. Eichmann, S. Preissler, R. Riek, and E. Deuerling. Cotranslational structure acquisition of nascent polypeptides monitored by NMR spectroscopy. *Proc. Natl. Acad. Sci. USA*, 107(20):9111–6, May 2010.

[9] L. D. Cabrita, C. M. Dobson, and J. Christodoulou. Protein folding on the ribosome. *Curr. Opin. Struc. Biol.*, 20(1):33–45, February 2010.

[10] L. D. Cabrita, C. M. Dobson, and J. Christodoulou. Early Nascent Chain Folding Events on the Ribosome. *Isr. J. Chem.*, 50(1):99–108, June 2010.

[11] L. D. Cabrita, A. M. E. Cassaignau, H. M. M. Launay, C. A. Waudby, T. Wlodarski, C. Camilloni, M.-E. Karyadi, A. L. Robertson, X. Wang, A. S. Wentink, L. S. Goodsell, C. A. Woolhead, M. Vendruscolo, C. M. Dobson, and J. Christodoulou. A structural ensemble of a ribosomeâĂŞnascent chain complex during cotranslational protein folding. *Nat. Struct. Mol. Biol.*, 23(4):278–285, 2016.

[12] S.-T. D. Hsu, L. D. Cabrita, P. Fucini, J. Christodoulou, and C. M. Dobson. Probing side-chain dynamics of a ribosome-bound nascent chain using methyl NMR spectroscopy. *J. Am. Chem. Soc.*, 131(24):8366–7, June 2009.

[13] C. Amero, P. Schanda, M. A. Durá, I. Ayala, D. Marion, B. Franzetti, B. Brutscher, and J. Boisbouvier. Fast two-dimensional NMR spectroscopy of high molecular weight protein assemblies. *J. Am. Chem. Soc.*, 131(10):3448–9, March 2009.

[14] S.-T. D. Hsu, P. Fucini, L. D. Cabrita, H. Launay, C. M. Dobson, and J. Christodoulou. Structure and dynamics of a ribosome-bound nascent chain by NMR spectroscopy. *Proc. Natl. Acad. Sci. USA*, 104(42):16516–21, October 2007.

[15] J. N. Onuchic, Z. Luthey-schulten, and P. G. Wolynes. Theory of Protein Folding : The Energy Landscape Perspective. *Annu. Rev. Phys. Chem.*, 48:545–600, 1997.

[16] C. M. Dobson, A. Šali, and M. Karplus. Protein folding: A perspective from theory and experiment. *Angew. Chem. Int. Ed.*, 37(7):868–893, 1998.

[17] R. Lill, E. Crooke, B. Guthrie, and W. Wickner. The "trigger factor cycle" includes ribosomes, presecretory proteins, and the plasma membrane. *Cell*, 54:1013–1018, 1988.

[18] M. P. Mayer and B. Bukau. Hsp70 chaperones: Cellular functions and molecular mechanism. *Cell. Mol. Life Sci.*, 62(6):670–684, 2005.

[19] A. C. Apetri and A. L. Horwich. Chaperonin chamber accelerates protein folding through passive action of preventing aggregation. *Proc. Natl. Acad. Sci. USA*, 105(45):17351–17355, 2008.

[20] A. L. Horwich, A. C. Apetri, and W. A. Fenton. The GroEL/GroES cis cavity as a passive anti-aggregation device. *FEBS lett.*, 583(16):2654–2662, 2009.

[21] D. Thirumalai and G. H. Lorimer. Chaperonin-mediated protein folding. *Annu. Rev. Biophys. Biomol. Struct.*, 30:245–269, 2001.

[22] D. Yang, X. Ye, and G. H. Lorimer. Symmetric GroEL:GroES2 complexes are the protein-folding functional form of the chaperonin nanomachine. *Proc. Natl. Acad. Sci. USA*, 110(46):E4298–E4305, 2013.

[23] G. Kramer, T. Rauch, W. Rist, S. Vorderwülbecke, H. Patzelt, A. Shulze-Specking, N. Ban, E. Deuerling, and B. Bukau. L23 protein functions as a chaperone docking site on the ribosome. *Nature*, 419:171–174, 2002.

[24] L. Ferbitz, T. Maier, H. Patzelt, B. Bukau, E. Deuerling, and N. Ban. Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature*, 431:590–596, 2004.

[25] T. Maier, L. Ferbitz, E. Deuerling, and N. Ban. A cradle for new proteins: trigger factor at the ribosome. *Curr. Opin. Struc. Biol.*, 15(2):204–12, April 2005.

[26] A. Hoffmann, B. Bukau, and G. Kramer. Structure and function of the molecular chaperone Trigger Factor. *Biochim. Biophys. Acta*, 1803:650–661, 2010.

[27] D. V. Fedyukina and S. Cavagnero. Protein folding at the exit tunnel. *Ann. Rev. Biophys.*, 40:337–359, 2011.

[28] E. P. O'Brien, J. Christodoulou, M. Vendruscolo, and C. M. Dobson. Trigger factor slows co-translational folding through kinetic trapping while sterically protecting the nascent chain from aberrant cytosolic interactions. *J. Am. Chem. Soc.*, 134(26):10920–32, July 2012.

[29] E. Martinez-Hackert and W. A. Hendrickson. Promiscuous Substrate Recognition in Folding and Assembly Activities of the Trigger Factor Chaperone. *Cell*, 138(5):923–934, 2009.

[30] M. Gamerdinger and E. Deuerling. Structural Basis for Protein Antiaggregation Activity of the Trigger Factor Chaperone. *Science*, 344(6184):590–591, 2014.

[31] A. Mashaghi, G. Kramer, P. Bechtluft, B. Zachmann-Brand, A. J. M. Driessen, B. Bukau, and S. J. Tans. Reshaping of the conformational search of a protein by the chaperone trigger factor. *Nature*, 500(7460):98–101, 2013.

[32] V. R. Agashe, S. Guha, H. C. Chang, P. Genevaux, M. Hayer-Hartl, M. Stemp, C. Georgopoulos, F. U. Hartl, and J. M. Barral. Function of trigger factor and DnaK in multidomain protein folding: Increase in yield at the expense of folding speed. *Cell*, 117:199–204, 2004.

[33] R. Gupta, S. K. Lakshmipathy, H. C. Chang, S. A. Etchells, and F. U. Hartl. Trigger factor lacking the PPIase domain can enhance the folding of eukaryotic multi-domain proteins in Escherichia coli. *FEBS Lett.*, 584(16):3620–3624, 2010.

[34] C. B. Anfinsen, E. Haber, M. Sela, and F. H. White Jr. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. USA*, 47(9):1309–1314, 1961.

[35] C. B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181(4096):223–30, July 1973.

[36] A. V. Nicola, W. Chen, and A. Helenius. Co-translational folding of an alphavirus capsid protein in the cytosol of living cells. *Nat. Cell. Biol.*, 1(6):341–345, 1999.

[37] J. Frydman, H. Erdjument-Bromage, P. Tempst, and F. U. Hartl. Co-translational domain folding as the structural basis for the rapid de novo folding of firefly luciferase. *Nat. Struct. Mol. Biol.*, 6(7):697–705, 1999.

[38] G. Zhang, M. Hubalewska, and Z. Ignatova. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat. Struct. Mol. Biol.*, 16(3):274–280, 2009.

[39] H. Nakatogawa and K. Ito. The Ribosomal Exit Tunnel Functions as a Discriminating Gate. *Cell*, 108:629–636, 2002.

[40] A. H. Elcock. Molecular Simulations of Cotranslational Protein Folding: Fragment Stabilities, Folding Cooperativity, and Trapping in the Ribosome. *PLoS Comput. Biol.*, 2(7):e98, 2006.

[41] E. P. O'Brien, S. T. D. Hsu, J. Christodoulou, M. Vendruscolo, and C. M. Dobson. Transient tertiary structure formation within the ribosome exit port. *J. Am. Chem. Soc.*, 132(47):16928–37, December 2010.

[42] J. Marino, G. Von Heijne, and R. Beckmann. Small protein domains fold inside the ribosome exit tunnel. *FEBS Letters*, 590(5):655–660, 2016.

[43] A. Kosolapov and C. Deutsch. Tertiary interactions within the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.*, 16(4):405–11, April 2009.

[44] M. R. Pool. A trans-membrane segment inside the ribosome exit tunnel triggers RAMP4 recruitment to the Sec61p translocase. *J. Cell. Biol.*, 185(5):889–902, jun 2009.

[45] S. Bhushan, T. Hoffmann, B. Seidelt, J. Frauenfeld, T. Mielke, O. Berninghausen, D. N. Wilson, and R. Beckmann. SecM-stalled ribosomes adopt an altered geometry at the peptidyl transferase center. *PLoS Biology*, 9(1):e1000581, 2011.

[46] A. A. Komar. A pause for thought along the co-translational folding pathway. *Trends Biochem. Sci.*, 34(1):16–24, January 2009.

[47] E. P. O'Brien, G. Stan, D. Thirumalai, and B. R. Brooks. Factors governing helix formation in peptides confined to carbon nanotubes. *Nano Lett.*, 8(11):3702–8, November 2008.

[48] H. Bremer and P. P. Dennis. (1996) Modulation of chemical composition and other parameters of the cell by growth rate. Neidhardt, et al. eds. Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology, 2nd ed. chapter 97, pp. 1559, Table 3.

[49] S. S. Jha and A. A. Komar. Using SecM arrest sequence as a tool to isolate ribosome bound polypeptides. *JoVE*, (64):1–5, January 2012.

[50] J. Christodoulou, G. Larsson, P. Fucini, S. R. Connell, T. A. Pertinhez, C. L. Hanson, C. Redfield, K. H. Nierhaus, C. V. Robinson, J. Schleucher, and C. M. Dobson. Heteronuclear NMR investigations of dynamic regions of intact Escherichia coli ribosomes. *Proc. Natl. Acad. Sci. USA*, 101(30):10949–54, July 2004.

[51] R. Augustyniak, F. Ferrage, C. Damblon, G. Bodenhausen, and P. Pelupessy. Efficient determination of diffusion coefficients by monitoring transport during recovery delays in NMR. *Chem. Commun.*, 48(43):5307–9, May 2012.

[52] S. Cai, C. Seu, Z. Kovacs, A. D. Sherry, and Y. Chen. Sensitivity Enhancement of Multidimensional NMR Experiments by Paramagnetic Relaxation Effects. *J. Am. Chem. Soc.*, 128(5):13474–13478, 2006.

[53] A. Deckert, C. A. Waudby, T. Wlodarski, A. S. Wentink, X. Wang, J. P. Kirkpatrick, J. F. S. Paton, C. Camilloni, P. Kukic, C. M. Dobson, M. Vendruscolo, L. D. Cabrita, and J. Christodoulou. Structural

characterization of the interaction of α-synuclein nascent chains with the ribosomal surface and trigger factor. *Proceedings of the National Academy of Sciences*, 113(18):5012–5017, 2016.

[54] I. S. Gabashvili, R. K. Agrawal, C. M. Spahn, R. A. Grassucci, D. I. Svergun, J. Frank, and P. Penczek. Solution structure of the E. coli 70S ribosome at 11.5 Å resolution. *Cell*, 100(5):537–49., 2000.

[55] P. Nissen, J. Hansen, N. Ban, P. B. Moore, and T. A. Steitz. The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. *Science*, 289:920–930, 2000.

[56] N. Ban, P. Nissen, J. Hansen, P. B. Moore, and T. A. Steitz. The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution. *Science*, 289(5481):905–920, August 2000.

[57] B. S. Schuwirth, M. A. Borovinskaya, C. W. Hau, W. Zhang, A. Vila-Sanjuro, J. M. Holton, and J. H. D. Cate. Structures of the Bacterial Ribosome at 3.5 Å Resolution. *Science*, 310:827–834, 2005.

[58] F. A. A. Mulder, L. Bouakaz, A. Lundell, M. Venkataramana, A. Liljas, M. Akke, and S. Sanyal. Conformation and Dynamics of Ribosomal Stalk Protein L12 in Solution and on the Ribosome. *Biochemistry*, 43:5930–5936, 2004.

[59] R. Riek, J. Fiaux, E. B. Bertelsen, A. L. Horwich, and K. Wuthrich. Solution NMR techniques for large molecular and supramolecular structures. *J. Am. Chem. Soc.*, 124(41):12144–12153, 2002.

[60] S. D. Hsu, L. D. Cabrita, P. Fucini, C. M. Dobson, and J. Christodoulou. Structure, dynamics and folding of an immunoglobulin domain of the gelation factor (ABP-120) from Dictyostelium discoideum. *J. Mol. Biol.*, 388(4):865–79, May 2009.

[61] P. Schanda, E. Kupce, and B. Brutscher. SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J. Biomol. NMR*, 33(4):199–211, December 2005.

[62] R. Rosenzweig and L. E. Kay. Bringing dynamic molecular machines into focus by methyl-TROSY NMR. *Annu. Rev. Biochem.*, 83:291–315, 2014.

[63] J. Cavanagh, W. J. Fairbrother, A. G. Palmer III, M. Rance, and N. J. Skelton. *Protein NMR Spectroscopy*. Academic Press, London, 2nd edition, 2007.

[64] O. Fisette, P. Lagüe, S. Gagné, and S. Morin. Synergistic applications of MD and NMR for the study of biological systems. *J. Biomed. Biotechnol.*, 2012:254208, January 2012.

[65] P. Neudecker, P. Lundström, and L. E. Kay. Relaxation dispersion NMR spectroscopy as a tool for detailed studies of protein folding. *Biophys. J.*, 96(6):2045–54, mar 2009.

[66] I. R. Kleckner and M. P. Foster. An introduction to NMR-based approaches for measuring protein dynamics. *BBA-Proteins Proteom.*, 1814(8):942–968, 2011.

[67] MATLAB and Statistics Toolbox Release 2012b, The MathWorks, Inc., Natick, Massachusetts, United States.

[68] P. Rovó, P. Stráner, A. Láng, I. Bartha, K. Huszár, L. Nyitray, and A. Perczel. Structural insights into the Trp-cage folding intermediate formation. *Chem. Eur. J.*, 19(8):2628–40, 2013.

[69] L. Qiu, S. A. Pabit, A. E. Roitberg, and S. J. Hagen. Smaller and faster: the 20-residue Trp-cage protein folds in 4 μs. *J. Am. Chem. Soc.*, 124(44):12952–3, December 2002.

[70] E. Lescop, P. Schanda, and B. Brutscher. A set of best triple-resonance experiments for time-optimized protein resonance assignment. *J. Mag. Res.*, 187(1):163–169, 2007.

[71] P. Schanda, H. Van Melckebeke, and B. Brutscher. Speeding up three-dimensional protein nmr experiments to a few minutes. *J. Am. Chem. Soc.*, 128(28):9042–9043, 2006.

[72] G. D. J. Phillies. Diffusion on a molecular scale as observed using PGSE NMR. *Concepts Magn. Reso. A*, 44(1):1–15, 2015.

[73] P. Schanda and B. Brutscher. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proetins on the time scale of seconds. *J. Am. Chem. Soc.*, 127(22):8014–8015, 2005.

[74] V. Tugarinov, P. M. Hwang, J. E. Ollerenshaw, and L. E. Kay. Cross-Correlated Relaxation Enhanced ^{1}H-^{13}C NMR Spectroscopy of Methyl Groups in Very High Molecular Weight Proteins and Protein Complexes. *J. Am. Chem. Soc.*, 125(34):10420–10428, 2003.

[75] J. E. Ollerenshaw, V. Tugarinov, and L. E. Kay. Methyl TROSY: explanation and experimental verification. *Magn. Reson. Chem.*, 41(10):843–852, October 2003.

[76] F. Bloch. Nuclear Induction. *Phys. Rev.*, 70(1946):460–74, 1946.

[77] H. M. McConnell. Reaction Rates by Nuclear Magnetic Resonance. *J. Chem. Phys.*, 28(3):430, 1958.

[78] D. F. Hansen and J. J. Led. Implications of using approximate Bloch–McConnell equations in NMR analyses of chemically exchanging systems: application to the electron self-exchange of plastocyanin. *J. Magn. Reson.*, 163(2):215–227, August 2003.

[79] R. Václav, H. Štěpánková, and J. Štěpánek. Analysis of NMR Spectra in Case of Temperature Dependent Chemical Exchange Between Two Unequally Populated Sites. *Concept. Magnetic Res. A*, 38A(3):117–127, 2011.

[80] U. Mayor, C. M. Johnson, V. Daggett, and A. R. Fersht. Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *PNAS*, 97(25):13518–13522, 2000.

[81] M. R. Bunagan, X. Yang, J. G. Saven, and F. Gai. Ultrafast folding of a computationally designed Trp-cage mutant: Trp2-cage. *J. Phys. Chem. B*, 110(8):3759–63, March 2006.

[82] T. Wang, Y. Zhu, and F. Gai. Folding of a three-helix bundle at the folding speed limit. *J. Phys. Chem.*, 108(12):3694–3697, 2004.

[83] Y. Zhu, D. O. V. Alonso, K. Maki, C. Huang, S. J. Lahr, V. Dagget, H. Roder, W. F. Degrado, and F. Gai. Ultrafast Folding of α_3D: A *De Novo* Designed Three-Helix Bundle Protein. *Proc. Natl. Acad. Sci. USA*, 100(26):15486–15491, 2003.

[84] J. Kubelka, T. K. Chiu, D. R. Davies, W. A. Eaton, and J. Hofrichter. Sub-microsecond protein folding. *J. Mol. Biol.*, 359(3):546–53, July 2006.

[85] S. Sato, T. L. Religa, and A. R. Fersht. Phi-analysis of the folding of the B domain of protein A using multiple optical probes. *J. Mol. Biol.*, 360(4):850–64, July 2006.

[86] M. Jäger, H. Nguyen, J. C. Crane, J. W. Kelly, and M. Gruebele. The folding mechanism of a β-sheet: the WW domain. *J. Mol. Biol.*, 311(2):373–93, August 2001.

[87] M. E. Rosenthal, A. D. Rojtman, and E. Frank. *Finegoldia magna* (formerly *Peptostreptococcus magnus*): an overlooked etiology for toxic shock syndrome? *Med. hypotheses*, 79(2):138–40, August 2012.

[88] J. F. Cramer, P. A. Nordberg, J. Hajdu, and S. Lejon. Crystal structure of a bacterial albumin-binding domain at 1.4 Å resolution. *FEBS Lett.*, 581(17):3178–82, July 2007.

[89] M. U. Johansson, M. de Château, L. Björk, S. Forsén, T. Drakenberg, and M. Wikström. The GA module, a mobile albumin-binding bacterial domain, adopts a three-helix-bundle structure. *FEBS Lett.*, 374:257–261, 1995.

[90] M. U. Johansson, M. de Château, M. Wikström, S. Forsén, T. Drakenberg, and L. Björk. Solution Structure of the Albumin-binding GA Module: A Versatile Bacterial Protein Domain. *J. Mol. Biol.*, 266:859–865, 1997.

[91] M. U. Johansson, I. M. Frick, P. J. Kraulis, S. Hober, M. Linhult, P. A. Nygren, M. Uhlén, L. Björck, T. Drakenberg, S. Forsén, and M. Wikström. Structure , Specificity , and Mode of Interaction for Bacterial Albumin-binding Modules. *J. Biol. Chem.*, 277(10):8114–8120, 2002.

[92] M. U. Johansson, H. Nilsson, J. Evenäs, S. Forsén, T. Drakenberg, L. Björk, and M. Wikström. Differences in Backbone Dynamics of Two Homologous Bacterial Albumin-binding Modules: Implications for Binding Specificity and Bacterial Adaptation. *J. Mol. Biol.*, 316:1083, 2002.

[93] Y. Zhu, X. Fu, T. Wang, A. Tamura, S. Takada, J. G. Saven, and F. Gai. Guiding the search for a protein's maximum rate of folding. *Chem. Phys.*, 307:99–109, 2004.

[94] S. Takada. Protein folding simulation with solvent-induced force field: folding pathway ensemble of three-helix-bundle proteins. *Proteins*, 42(1):85–98, January 2001.

[95] H. Lei and Y. Duan. Ab Initio Folding of Albumin Binding Domain from All-Atom Molecular Dynamics Simulation. *J. Phys. Chem. B*, 111:5458–5463, 2007.

[96] T. Cellmer, E. R. Henry, J. Kubelka, J. Hofrichter, and W. A. Eaton. Relaxation rate for an ultrafast folding protein is independent of chemical denaturant concentration. *J. Am. Chem. Soc.*, 129(47):14564–14565, 2007.

[97] S. H. Brewer, D. M. Vu, Y. Tang, Y. Li, S. Franzen, D. P. Raleigh, and R. B. Dyer. Effect of modulating unfolded state structure on the folding kinetics of the villin headpiece subdomain. *Proc. Natl. Acad. Sci. USA*, 102(46):16662–16667, 2005.

[98] J. Kubelka, W. A. Eaton, and J. Hofrichter. Experimental tests of villin subdomain folding simulations. *J. Mol. Biol.*, 329(4):625–630, 2003.

[99] M. Wang, Y. Tang, S. Sato, L. Vugmeyster, C. J. McKnight, and D. P. Raleigh. Dynamic NMR line-shape analysis demonstrates that the villin headpiece subdomain folds on the microsecond time scale. *J. Am. Chem. Soc.*, 125(20):6032–6033, 2003.

[100] P. V. Banushkina and S. V. Krivov. High-resolution free-energy landscape analysis of alpha-helical protein folding: HP35 and its double mutant. *J. Chem. Theory Comput.*, 9(12):5257–5266, 2013.

[101] J. Kubelka, J. Hofrichter, and W. A. Eaton. The protein folding 'speed limit'. *Curr. Opin. Struc. Biol.*, 14(1):76–88, February 2004.

[102] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw. How fast-folding proteins fold. *Science (New York, N.Y.)*, 334(6055):517–20, October 2011.

[103] T. K. Chiu, J. Kubelka, R. Herbst-Irmer, W. a. Eaton, J. Hofrichter, and D. R. Davies. High-resolution x-ray crystal structures of the villin headpiece subdomain, an ultrafast folding protein. *Proc. Natl. Acad. Sci. USA*, 102(21):7517–7522, 2005.

[104] L. Wickstrom, Y. Bi, V. Hornak, D. P. Raleigh, and C. Simmerling. Reconciling the solution and X-ray structures of the villin headpiece helical subdomain: Molecular dynamics simulations and double mutant cycles reveal a stabilizing cation-π interaction. *Biochemistry*, 46(12):3624–3634, 2007.

[105] M. R. Bunagan, J. Gao, J. W. Kelly, and F. Gai. Probing the folding transition state structure of the villin headpiece subdomain via side chain and backbone mutagenesis. *J. Am. Chem. Soc.*, 131(21):7470–7476, 2009.

[106] L. E. Packer, B. Song, D. P. Raleigh, and C. J. McKnight. Competition between intradomain and interdomain interactions: A buried salt bridge is essential for villin headpiece folding and actin binding. *Biochemistry*, 50(18):3706–3712, 2011.

[107] A. L. Serrano, O. Bilsel, and F. Gai. Native state conformational heterogeneity of HP35 revealed by time-resolved FRET. *J. Phys. Chem. B*, 116(35):10631–8, sep 2012.

[108] Y. Bi, Y. Tang, D. P. Raleigh, and J. H. Cho. Efficient high level expression of peptides and proteins as fusion proteins with the N-terminal domain of L9: Application to the villin headpiece helical subdomain. *Protein Expres. Purif.*, 47(1):234–240, 2006.

[109] W. Humphrey, A. Dalke, and K. Schulten. VMD - Visual Molecuar Dynamics. *J. Molec. Graphics.*, 14(1):33–38, 1996.

[110] T. Arakawa and K. Tsumoto. The effects of arginine on refolding of aggregated proteins: not facilitate refolding, but suppress aggregation. *Biochem. Bipohys. Res. Commun.*, 304(1):148–152, 2003.

[111] C. P. Schneider and B. L. Trout. Investigation of cosolute-protein preferential interaction coefficients: new insight into the mechanism by which arginine inhibits aggregation. *J. Phys. Chem. B*, 113:2050–2058, 2009.

[112] L. Ito, K. Shiraki, T. Matsuura, M. Okumura, K. Hasegawa, S. Baba, H. Yamaguchi, and T. Kumasaka. High-resolution X-ray analysis reveals binding of arginine to aromatic residues of lysozyme surface: implication of suppression of protein aggregation by arginine. *Protein Eng. Des. Sel.*, 24:269–274, 2011.

[113] R. Ariki, T. Hirano, and K. Shiraki. Arginine increases the solubility of alkyl gallates through interaction with the aromatic ring. *J. Biochem.*, 149:389–394, 2011.

[114] A. Hirano, T. Kameda, D. Shinozaki, T. Arakawa, and K. Shiraki. Molecular dynamics simulation of the arginine-assisted solubilization of caffeic acid: intervention in the interaction. *J. Phys. Chem.*, 117:7518–7527, 2013.

[115] R. Harada, N. Tochio, T. Kigawa, Y. Sugita, and M. Feig. Reduced Native State Stability in Crowded Cellular Environment Due to Protein âĹŠ Protein Interactions. *J. Am. Chem. Soc.*, 135:3696–701, 2013.

[116] A. Murakami, H. Nakatogawa, and K. Ito. Translation arrest of SecM is essential for the basal and regulated expression of SecA. *PNAS*, 101(33):12330–5, August 2004.

[117] P. Sensi, P. Margalith, and M. T. Timbal. Rifomycin, a new antibiotic-prelimanary report. *Farmaco Ed. Sci.*, 14:146–147, 1959.

[118] E. O. Stejskal and J. E. Tanner. Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient. *J. Chem. Phys.*, 42(1):288–292, 1965.

[119] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, and E. D. Laue. The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins*, 59(4):687–696, 2005.

[120] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, and N. A. Baker. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.*, 32(Web Server issue):W665–7, July 2004.

[121] Y. Dehouck, J. M. Kwasigroch, D. Gilis, and M. Rooman. PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC Bioinformatics*, 12(1):151, January 2011.

[122] J. Iwahara, C. Tang, and G. M. Clore. Practical aspects of ^{1}H transverse paramagnetic relaxation enhancement measurements on macromolecules. *J. Magn. Reson.*, 184(2):185–95, February 2007.

[123] J. M. Rogers, A. Steward, and J. Clarke. Folding and Binding of an Intrinsically Disordered Protein: Fast, but not 'Diffusion-Limited'. *J. Am. Chem. Soc.*, 135(4):1415–22, January 2013.

[124] J. F. Eichler, J. C. Cramer, K. L. Kirk, and J. G. Bann. Biosynthetic incorporation of fluorohistidine into proteins in *E. coli*: A new probe of macromolecular structure. *ChemBioChem*, 6(12):2170–2173, 2005.

[125] M. D. Vaughan, P. Cleve, V. Robinson, H. S. Duewel, and J. F. Honek. Difluoromethionine as a novel ^{19}F NMR structural probe for internal amino acid packing in proteins. *J. Am. Chem. Soc.*, 121(37):8475–8478, 1999.

[126] C. Li, E. A. Lutz, K. M. Slade, R. A. S. Ruf, G. F. Wang, and G. J. Pielak. ^{19}F NMR studies of α-synuclein conformation and fibrillation. *Biochemistry*, 48(36):8578–84, September 2009.

[127] G. Marsh, E. Neil, and Y. Suzuki. Using ^{19}F NMR to Probe Biological Interactions of Proteins and Peptides. *Chemical biology*, 9:1242–1250, 2014.

[128] R. Curtis-Marof, D. Doko, M. L. Rowe, K. L. Richards, R. A. Williamson, and M. J. Howard. ^{19}F NMR spectroscopy monitors ligand binding to recombinantly fluorine-labelled b'x from human protein disulphide isomerase (hPDI). *Org. Biomol. Chem.*, 12(23):3808–12, jun 2014.

[129] F. Khan, I. Kuprov, T. D. Craggs, P. J. Hore, and S. E. Jackson. ^{19}F NMR studies of the native and denatured states of green fluorescent protein. *J. Am. Chem. Soc.*, 128(33):10729–10737, 2006.

[130] J. L. Kitevski-Leblanc, J. Hoang, W. Thach, S. T. Larda, and R. S. Prosser. ^{19}F NMR studies of a desolvated near-native protein folding intermediate. *Biochemistry*, 52(34):5780–9, August 2013.

[131] C. Li, G.-F. Wang, Y. Wang, R. Creager-Allen, E. A. Lutz, H. Scronce, K. M. Slade, R. A. S. Ruf, R. A. Mehl, and G. J. Pielak. Protein ^{19}F NMR in Escherichia coli. *J. Am. Chem. Soc.*, 132(1):321–327, 2010.

[132] Y. Ye, X. Liu, Z. Zhang, Q. Wu, B. Jiang, L. Jiang, X. Zhang, M. Liu, G. J. Pielak, and C. Li. ^{19}F NMR spectroscopy as a probe of cytoplasmic viscosity and weak protein interactions in living cells. *Chem. Eur. J.*, 19:12705–10, 2013.

[133] M. Veronesi, F. Giacomina, E. Romeo, B. Castellani, G. Ottonello, C. Lambruschini, G. Garau, R. Scarpelli, T. Bandiera, D. Piomelli, and C. Dalvit. Fluorine NMR-based assay in living mammalian cells. *Anal. Biochem.*, 495(December):1–8, 2015.

[134] E. Luchinat and L. Banci. A unique tool for cellular structural biology: In-cell NMR. *J. Biol. Chem.*, 291(8):3776–3784, 2016.

[135] C. Minks, S. Alefelder, L. Moroder, R. Huber, and N. Budisa. Towards New Protein Engineering: In Vivo Building and Folding of Protein Shuttles for Drug Delivery and Targeting by the Selective Pressure Incorporation (SPI) Method. *Tetrahedron*, 56(48):9431–9442, 2000.

[136] Y. Tang and D. A. Tirrell. Biosynthesis of a highly stable coiled-coil protein containing hexafluoroleucine in an engineered bacterial host [13]. *J. Am. Chem. Soc.*, 123(44):11089–11090, 2001.

[137] J. W. Peng. Cross-Correlated ^{19}F Relaxation Measurements for the Study of Fluorinated Ligand-Receptor Interactions. *J. Magn. Reson.*, 153(1):32–47, 2001.

[138] H. Duewel, E. Daub, V. Robinson, and J. F. Honek. Incorporation of trifluoromethionine into a phage lysozyme: Implications and a new marker for use in protein ^{19}F NMR. *Biochemistry*, 36(11):3404–3416, 1997.

[139] H. S. Duewel, E. Daub, V. Robinson, and J. F. Honek. Elucidation of solvent exposure, side-chain reactivity, and steric demands of the trifluoromethionine residue in a recombinant protein. *Biochemistry*, 40(44):13167–13176, 2001.

[140] W. Kim, A. George, M. Evans, and V. P. Conticello. Cotranslational incorporation of a structurally diverse series of proline analogues in an Escherichia coli expression system. *ChemBioChem*, 5(7):928–936, 2004.

[141] Q. S. Zhang, L. Shen, E. D. Wang, and Y. L. Wang. Biosynthesis and characterization of 4-fluorotryptophan-labeled Escherichia coli arginyl-tRNA synthetase. *J. Protein Chem.*, 18(2):187–192, 1999.

[142] C. A. Waudby, M. D. Mantle, L. D. Cabrita, L. F. Gladden, C. M. Dobson, and J. Christodoulou. Rapid Distinction of Intracellular and Extracellular Proteins Using NMR Diffusion Measurements. *J. Am. Chem. Soc.*, 134:11312–11315, 2012.

[143] K. Knoops, G. Schoehn, and C. Schaffitzel. Cryo-electron microscopy of ribosomal complexes in cotranslational folding, targeting, and translocation. *WIREs RNA*, 3(3):429–41, 2012.

[144] S. A. Weinreis, J. P. Ellis, and S. Cavagnero. Dynamic fluorescence depolarization: a powerful tool to explore protein folding on the ribosome. *Methods*, 52(1):57–73, September 2010.

[145] T. Y. Kim, T. Schlieter, S. Haase, and U. Alexiev. Activation and molecular recognition of the GPCR rhodopsin - Insights from time-resolved fluorescence depolarisation and single molecule experiments. *Eur. J. Cell Biol.*, 91(4):300–310, 2012.

[146] A. Ben-shem, L. Jenner, G. Yusupova, and M. Yusupov. Crystal Structure of the Eukaryotic Ribosome. *Science*, 330:1203–1209, 2010.

[147] Y. Zhang, T. Wolfle, and S. Rospert. Interaction of nascent chains with the ribosomal tunnel proteins Rpl4, Rpl17, and Rpl39 of Saccharomyces cerevisiae. *J. Biol. Chem.*, 288(47):33697–33707, 2013.

[148] M. T. Lin, L. J. Sperling, H. L. Frericks Schmidt, M. Tang, R. I. Samoilova, T. Kumasaka, T. Iwasaki, S. A. Dikanov, C. M. Rienstra, and R. B. Gennis. A rapid and robust method for selective isotope labeling of proteins. *Methods*, 55(4):370–378, 2011.

[149] T. Iwasaki, R. Fukazawa, Y. Miyajima-Nakano, A. Baldansuren, S. Matsushita, M. T. Lin, R. B. Gennis, K. Hasegawa, T. Kumasaka, and S. A. Dikanov. Dissection of hydrogen bond interaction network around an iron-sulfur cluster by site-specific isotope labeling of hyperthermophilic archaeal Rieske-type ferredoxin. *J. Am. Chem. Soc.*, 134:19731–19738, 2012.

[150] M. T. Lin, R. Fukazawa, Y. Miyajima-Nakano, S. Matsushita, S. K. Choi, T. Iwasaki, and R. B. Gennis. Escherichia coli auxotroph host strains for amino acid-selective isotope labeling of recombinant proteins. *Methods Enzymol.*, 565:45–66, 2015.

[151] M. A. Motin, T. K. Biswas, and E. M. Huque. Volumetric and viscometric studies on an aqueous urea solution. *Phys. Chem. Liq.*, 40(5):593–605, 2002.

[152] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, 6(3):277–293, 1995.

[153] P. R. Bevington and D. K. Robinson. *Data Reduction And Error Analysis for the Physical Sciences*. McGraw-Hill, New York, 3rd ed edition, 2003.

[154] A. Bax, G. M. Clore, and A. Gronenborn. ^{1}H ^{1}H correlation via isotropic mixing of ^{13}C magnetization, a new three-dimensional approach for assigning ^{1}H and ^{13}C spectra of ^{13}C-enriched proteins. *J. Magn. Reson.*, 88:425–431, 1990.